

ADDITIONAL REFERENCES FOR CHAPTER 32

- BARRON, E. S. G. AND SINGER, T. P. *J. Biol. Chem.* 1945, 257, 221, 241
- BURTON, K. AND KRESS, H. A. *Biochem. J.* 1953, 54, 94
- GILLISPIE, R. J., MATY, G. A. AND VERNON, C. A. *Nature* 1953, 171, 1147
- HOGESCOM, G. H. *J. Biol. Chem.* 1949, 177, 547
- HORECKER, B. L. *Federation Proceedings* 1949, 8, 207
- HORECKER, B. L. *J. Biol. Chem.* 1950, 183, 539
- HORECKER, B. L. AND KOEHLING, A. J. *J. Biol. Chem.* 1946, 163, 11
- KEHLIN, D. AND HARTREE, E. F. *Proc. Roy. Soc. B.* 1959, 127, 167
- KEHLIN, D. AND HARTREE, E. F. *Proc. Roy. Soc. B.* 1945, 39, 293
- KEHLIN, D. AND HARTREE, E. F. *Proc. Roy. Soc. B.* 1947, 41, 500
- KEHLIN, D. AND HARTREE, E. F. *Biochem. J.* 1949, 44, 205
- MAYER, H. R. AND ELVING, D. G. *J. Biol. Chem.* 1952, 210, 165
- RAVE, T. W. AND LINDNER, A. L. *J. Biol. Chem.* 1952, 194, 119
- RAND, L. J. AND DEBUS, B. G. *J. Amer. Chem. Soc.* 1953, 75, 1261
- THEORELL, H. AND ANDERSSON, A. *J. Amer. Chem. Soc.* 1941, 63, 157
- HANES, C. S., CONNELL, G. E. AND DIXON, G. H. *Thermodynamic and thermodynamic reactions of metabolism* Vol. II, Baltimore, The Johns Hopkins Press 1952
- HEARON, J. Z. *Thermodynamic principle as applied to the analysis of a logical system* *Federation Proceedings* 1951, 10, 602
- HEARON, J. Z. *Rate behavior of metabolic systems* *Physiol. Rev.* 1952, 32, 499
- HEFFERMAN, L. *Reversible inactivations of certain catalytic enzymes* *Physiol. Rev.* 1957, 17, 474
- KEHLIN, D. *Cytochrome and intracellular respiration* *Enzymology* 1953, 2, 239
- LEACH, S. J. *The mechanism of enzyme oxidation-reduction* *Advances in Enzymology* 1954, 15, 153
- LEACH, R. S. *General Biology and Pathology of Organisms* University of Chicago Press, 1945
- LEPMANN, F. *Metabolic generation and utilization of phosphate bond energy* *Advances in Enzymology* 1941, 1, 100
- LEPMANN, F. *Aerobic phosphate* *Advances in Enzymology* 1946, 6, 231
- LEON, LEO. *The Biological Basis of Individuality* C. C. Thomas, Springfield, Ill. 1945
- LYNN, F. *Functional groups of coenzyme A and its metabolic relations* *Federation Proceedings* 1953, 12, 683
- LYNN, F. *The participation of coenzyme A in the oxidation of fat* *Nature* 1954, 174, 962
- MAYER, H. R. *Role of coenzyme A in fatty acid metabolism* *Federation Proceedings* 1953, 12, 692
- MAYER, H. R. AND GREEN, D. E. *Metabolic pathways and electron transport* *Science* 1954, 120, 7
- NOVILLI, G. D. *Enzymatic synthesis and structure of coenzyme A* *Federation Proceedings* 1953, 12, 675
- OSHOA, S. *Enzymic mechanisms of the citric acid cycle* *Advances in Enzymology* 1954, 15, 133
- OGSTON, A. G. AND SMITH, O. *Somethermodynamic and kinetic aspects of metabolic phosphorylation* *Physiol. Rev.* 1948, 28, 233
- OFFENHEIMER, C. AND STERN, K. G. *Biological Oxidation* Nordmann, New York, 1959
- PART, K. G. *The iron-containing enzymes* *Cytochromes* Summer J. B. and Myrback K. The Enzymes Vol. II Part I. Academic Press New York, 1951
- SINGER, T. P. AND KEARNEY, E. B. *Chemical metabolism and scope of action of the pyridine nucleotide coenzymes* *Advances in Enzymology* 1954, 15, 79
- SLATER, E. C. *Biological oxidation* *Ann. Rev. Biochem.* 1953, 22, 17
- THEORELL, H. *Heat-linked groups and mode of action of some hemoproteins* *Advances in Enzymology* 1947, 7, 265
- THEORELL, H. *Fluorine-containing enzymes* Summer J. B. and Myrback, K. The Enzymes Vol. II, Part I. Academic Press, New York, 1951
- THEORELL, H. *The iron-containing enzymes* *Cytochromes* Summer J. B. and Myrback, K. The Enzymes Vol. II, Part I. Academic Press, New York, 1951
- WARBURG, O. *The respiration enzyme* *Science* 1925, 61, 57, ibid 1928, 63, 437
- WARBURG, O. *Heavy Metal Prothetetic Groups* Clarendon Press, Oxford, 1949
- Monographs and Reviews*
- BARRON, E. S. G. *Mechanisms of carbohydrate metabolism* *Advances in Enzymology*, 1943, 3, 1-9
- BARRON, E. S. G. *The application of biological oxidation-reduction reaction systems to the study of cellular respiration* *Biological Symposia* 1943, 10, 27
- BURTON, T. *Thermodynamik und Enzyme* *Ergebe. Enzymforsch.* 1953, 4, 68
- DAWSON, C. F. AND TAPLIN, W. B. *The copper oxidases* Summer, J. B. and Myrback, K. The Enzymes Vol. II, Part I. Academic Press, New York, 1951
- DRABIN, D. L. *Distribution and metabolic aspects of derivatives of iron protoporphyrin* *Federation Proceedings* 1948, 7, 483
- DRABIN, D. L. *Metabolism of the heme chromoproteins* *Physiol. Rev.* 1951, 31, 345
- ELLIOTT, K. A. C. *Biological oxidations and reductions* *Ann. Rev. Biochem.* 1946, 15, 1
- FISCHER, F. G. *Nucleonuclear Überträger biologischer Oxydo-Reduktionen und ihre Potentiale* *Ergebe. Enzymforsch.* 1959, 8, 186
- GLUTATHIONE. *A symposium* (Editors: Colovick et al.) Academic Press, New York, 1954
- GRANICK, S. AND GILDER, H. *Distribution, structure and properties of the tetrapyrroles* *Advances in Enzymology* 1947, 7, 265
- GREEN, D. E. *Mechanism of Biological Oxidations* Cambridge University Press, 1940
- GREEN, D. E. *Fatty acid oxidation in soluble enzyme systems of animal tissues* *Biological Reviews* 1954, 29, 330
- GRUBBS, I. C. *Group-transfer and acyl-generating functions of lipoc acid derivatives* Summer, J. B. and Myrback, K. The Enzymes Vol. II, Part I. Academic Press, 1953
- GRUBBS, I. C. *Oxidative and transfer reactions of lipoc acid* *Federation Proceedings* 1954, 13, 715
- HANES, C. S. *The formation of peroxides in enzymic reactions* *British Medical Bulletin* 1953, 9, 131

PREFACE TO THE SIXTH EDITION

In the preparation of this edition the book has undergone the most extensive revision since it was first published. Scarcely a page has escaped some emendation, deletion or addition. Many figures have been redrawn and others added. The bulk of the book has been reduced through the elimination of older material where the space could be occupied more advantageously by accounts of recent advances in the subject, and by a reduction in the size of some of the illustrations and tables.

As a convenience to readers when referring from one part of the text to another, or to relevant articles in the literature, chapter numbers have been printed at the heads of the pages. The index has been completely redone and greatly enlarged.

The following paragraph is reproduced from the preface of the last edition, as it applies equally well to this one.

"The authors have received many letters relating to some or other statement in the book, the criticisms, comments and suggestions which have been so kindly offered are most cordially welcomed. These letters seem to display an interest in the welfare of the book which has been a source of satisfaction and pleasure to us, they have been of much assistance in the preparation of this edition. We are also most grateful to many of our friends and colleagues who have sent us reprints of their articles, it is hoped that even greater numbers will be received for use in the preparation of the next edition. Since one of the authors has transferred to the staff of another university, and in order that it may be known to whom letters or reprints should be addressed, it has been thought advisable that the parts of the text originally written by this author and revised by him in succeeding editions should be shown. His responsibility has, therefore, been indicated by initials at the heads of sections."

The most sincere thanks are again extended to Professor Arthur M. Wynne for the revision of his section on Biological Oxidations—a rapidly advancing subject of great complexity, and one which only a specialist in this field could present with clarity.

N B T

PREFACE TO FIRST EDITION

Physiology is a science in its own right and the laboratory worker who pursues his researches quite detached from medical problems need offer no apology for his academic outlook. Indeed some of the most valuable contributions to medical science have been the outcome of laboratory studies whose applications could not have been foreseen. Nevertheless, we feel that the teacher of physiology in a medical school owes it to his students, whose ultimate interest it must be conceded is in the diagnosis and treatment of disease, to emphasize those aspects of the subject which will throw light upon disorders of function. The physiologist can in this way play a part in giving the student and practitioner a vantage point from which he may gain a rational view of pathological processes.

We have endeavored to write a book which will serve to link the laboratory and the clinic, and which will therefore promote continuity of physiological teaching throughout the pre-clinical and clinical years of the under-graduate course. It is also hoped that when the principles underlying diseased states are pointed out to the medical student, and he is shown how a knowledge of such principles aids in the interpretation of symptoms or in directing treatment, he will take a keener interest in physiological studies. When such studies are restricted to the classical aspects of the subject, apparently remote from clinical application, the student is likely to regard them only as a task which his teachers in their inscrutable wisdom have condemned him to perform. Too often he gains the idea, from such a course, that physiology is of very limited utility and comes to believe that, having once passed into the clinical years, most of what he has "crammed" for examination purposes may be forgotten without detriment to his more purely medical studies. Unfortunately, he does not always realize at this stage in his education how great has been the part which physiological discoveries have played in the progress of medicine, and that the practice of today has evolved from the "theories" of yesterday.

Many physiological problems can be approached only through animal experimentation. Advances in many fields, most notably in those of carbohydrate metabolism, nutrition, and endocrinology, bear witness to the fertility of this method of research. On the other hand, many problems can be elucidated only by observations upon man, and physiology has gained much from clinical research. The normal human subject as an experimental animal possesses unique advantages for many types of investigation, and in disease, nature produces abnormalities of structure and function which the physiological laboratory can imitate only in the crudest way. Within recent years the clinical physiologist, fully realizing these advantages and the opportunities afforded by the hospital wards, has contributed very largely to physiological knowledge. In many instances, clinical research has not only revealed the true nature of the underlying process in disease, but has cast a light into some dark corner of physiology as well, several examples of clinical investigation which have pointed the way to the physiologist could be cited. In the last century, knowledge of the processes of disease was sought mainly in studies of morbid *anatomy*, biochemistry was in its infancy and many of the procedures now commonly employed for the investigation of the human subject had not been devised. Today, the student of scientific medicine is directing his attention more and more to the study of morbid *physiology* in his efforts to solve clinical problems. This newer outlook has borne fruit in many fields. It has had the beneficent result of drawing the clinic and the physiological and bio-

chemical laboratories onto common ground from which it has often been possible to launch a joint attack upon disease. We feel that this modern trend in the field of research should be reflected in the teaching of medical students, and have therefore given greater prominence to clinical aspects of the subject than is usual in physiological texts.

In order to understand the function of an organ it is usually essential to have a knowledge of its structure. For this reason we have followed the plan of preceding the account of the physiology of a part by a short description of its morphology and, in many instances, of its nerve and blood supply. The architecture and functions of the central nervous system are so intimately related that some space has been devoted to a description of the more important fiber tracts and grey masses of the cerebrum, cerebellum and spinal cord.

We wish to thank our colleagues in physiology, biochemistry and anatomy whom we have drawn upon on so many occasions for information and advice, without their generous help the undertaking would have been an almost impossible one. We are also deeply grateful for the unstinted assistance which we have received from our friends on the clinical staff, several of whom have read parts of the text in manuscript or in proof. We wish especially to acknowledge our indebtedness to Professor A. M. Wynne, who has written the section on the oxidizing systems of living cells, to Dr. J. K. W. Ferguson for his collaboration in the preparation of Chapter 33, and to Professor C. B. Weld and Dr. E. T. Waters whose stimulating criticisms and sound counsel have been invaluable.

Finally, we wish to thank our secretaries, Miss Mabel Cory and Miss Dudley Martin, who have spent so many tedious hours in preparing the manuscript for the press, in checking the references and in compiling the index.

October 15, 1936

C H B
N B T

CONTENTS

SECTION I THE BLOOD AND LYMPH

CHAPTER 1

THE PHYSIOLOGICAL PROPERTIES, PHYSICAL CHARACTERS AND COMPOSITION OF THE BLOOD	1
--	---

CHAPTER 2

THE RED CELLS OR ERYTHROCYTES	9
-------------------------------	---

CHAPTER 3

THE BLOOD VOLUME BODY WATER WATER BALANCE	17
---	----

CHAPTER 4

THE LYMPH AND TISSUE FLUIDS	33
-----------------------------	----

CHAPTER 5

TRANSFUSION THE BLOOD GROUPS	41
------------------------------	----

CHAPTER 6

HEMOCLOBIN	53
------------	----

CHAPTER 7

HEMOLYSIS AND SUSPENSION STABILITY OF THE BLOOD	63
---	----

CHAPTER 8

THE SPLEEN, THE LIFE OF THE RED CELL, THE REGENERATION OF BLOOD, IRON METABOLISM	69
--	----

CHAPTER 9

THE ANEMIAS	80
-------------	----

CHAPTER 10

THE WHITE BLOOD CORPUSCLES OR LEUKOCYTES —THE PLATELETS	92
---	----

CHAPTER 11

THE ORIGIN OF THE BLOOD CELLS—HEMATOPOIESIS	103
---	-----

CHAPTER 12

THE COAGULATION (CLOTTING) OF BLOOD	113
-------------------------------------	-----

CHAPTER 13

THE MECHANISMS REGULATING THE REACTION OF THE BODY FLUIDS	126
---	-----

SECTION II THE CIRCULATION OF THE BLOOD

CHAPTER 14

THE DYNAMICS OF THE CIRCULATION	138
---------------------------------	-----

CHAPTER 15

THE ARTERIAL BLOOD PRESSURE	146
-----------------------------	-----

CHAPTER 16

THE BLOOD PRESSURE IN THE HUMAN SUBJECT THE VENOUS BLOOD FLOW :	152
---	-----

CHAPTER 17	
THE VELOCITY OF THE BLOOD	175
CHAPTER 18	
THE ARTERIAL PULSE	184
CHAPTER 19	
THE PHYSIOLOGY OF CARDIAC MUSCLE PERFUSION OF THE ISOLATED HEART	188
CHAPTER 20	
THE ORIGIN AND MODE OF TRANSMISSION OF THE HEART BEAT	196
CHAPTER 21	
THE ARCHITECTURE OF THE HEART, MOVEMENTS OF THE HEART, THE CARDIAC CYCLE	203
CHAPTER 22	
THE MOVEMENTS OF THE HEART VALVES THE HEART SOUNDS	210
CHAPTER 23	
ELECTROCARDIOGRAPHY THE VENOUS PULSE	217
CHAPTER 24	
DISORDERS OF THE HEART BEAT AND THEIR INVESTIGATION BY GRAPHIC METHODS	227
CHAPTER 25	
THE REGULATION OF THE HEART'S ACTION, HEART FAILURE	241
CHAPTER 26	
THE OUTPUT OF THE HEART	263
CHAPTER 27	
THE CONTROL OF THE BLOOD VESSELS THE VASOMOTOR MECHANISMS PERIPHERAL VASCULAR DISORDERS SURGICAL SHOCK	273
CHAPTER 28	
SPECIAL FEATURES OF THE CIRCULATION IN DIFFERENT REGIONS	308
SECTION III RESPIRATION	
CHAPTER 29	
INTRODUCTION THE MECHANICS OF RESPIRATION	343
CHAPTER 30	
THE AIR OF THE LUNGS	360
CHAPTER 31	
THE PHYSICAL PRINCIPLES GOVERNING THE RESPIRATORY EXCHANGES	365
CHAPTER 32	
THE TRANSPORT AND DELIVERY OF OXYGEN TO THE TISSUES, INTRACELLULAR OXIDA- TIONS AND ENERGY TRANSFER, THE CARRIAGE OF CARBON DIOXIDE	371
CHAPTER 33	
THE CONTROL OF RESPIRATION, PERIODIC RESPIRATION, DISPNEA	400
CHAPTER 34	
ANOXIA (MORE CORRECTLY HYPOXIA)	419

SECTION IV THE EXCRETION OF URINE

CHAPTER 35

URINE FORMATION, THE STRUCTURE OF THE KIDNEY, THEORIES OF RENAL FUNCTION, VOLUME AND COMPOSITION OF THE URINE	440
--	-----

CHAPTER 36

THE PATHOLOGICAL PHYSIOLOGY OF KIDNEY DISEASE MICTURITION	467
---	-----

SECTION V DIGESTION

CHAPTER 37

THE SALIVARY GLANDS AND THE SECRETION OF SALIVA	487
---	-----

CHAPTER 38

GASTRIC DIGESTION	496
-------------------	-----

CHAPTER 39

GASTRIC DIGESTION (CONTINUED) THE SECRETION OF GASTRIC JUICE	505
--	-----

CHAPTER 40

DIGESTION IN THE INTESTINE	527
----------------------------	-----

CHAPTER 41

✓ THE GALL-BLADDER AND BILE DUCTS	551
-----------------------------------	-----

CHAPTER 42

THE MOVEMENTS OF THE ALIMENTARY CANAL	557
---------------------------------------	-----

CHAPTER 43

MOVEMENTS OF THE ALIMENTARY CANAL (CONTINUED) ABSORPTION FROM THE IN- TESTINES, THE FORMATION OF FECES, CONSTIPATION, INTESTINAL OBSTRU- TION	578
---	-----

CHAPTER 44

VISCERAL SENSATIONS	597
---------------------	-----

SECTION VI METABOLISM AND NUTRITION

CHAPTER 45

GENERAL METABOLISM	606
--------------------	-----

CHAPTER 46

PROTEIN METABOLISM	622
--------------------	-----

CHAPTER 47

PROTEIN METABOLISM (CONTINUED)	643
--------------------------------	-----

CHAPTER 48

THE METABOLISM OF THE PURINE AND PYRIMIDINE BASES	651
---	-----

CHAPTER 49

CARBOHYDRATE METABOLISM	658
-------------------------	-----

CHAPTER 50

FAT METABOLISM	687
----------------	-----

CHAPTER 51	
THE METABOLISM IN STARVATION, SEMISTARVATION AND OBESITY	707
CHAPTER 52	
METABOLISM IN MUSCULAR ACTIVITY	714
CHAPTER 53	
THE BODY TEMPERATURE HEAT BALANCE	730
CHAPTER 54	
THE VITAMINS	743
CHAPTER 55	
THE VITAMINS (CONTINUED)	761
CHAPTER 56	
DIETARY REQUIREMENTS	775
SECTION VII THE DUCTLESS GLANDS OR ENDOCRINES	
CHAPTER 57	
INTRODUCTION THE PITUITARY BODY	782
CHAPTER 58	
THE THYROID GLAND	806
CHAPTER 59	
THE ADRENAL GLANDS (SUPRARENAL CAPSULES)	827
CHAPTER 60	
THE PARATHYROID GLANDS AND CALCIUM METABOLISM	849
CHAPTER 61	
THE ENDOCRINE ORGANS OF SEX THE THYMUS AND PINEAL GLANDS	871
SECTION VIII THE NERVOUS SYSTEM	
CHAPTER 62	
INTRODUCTORY THE PHYSIOLOGICAL PROPERTIES OF NERVE	910
CHAPTER 63	
THE REFLEX ARC. RECEPTOR ORGANS CUTANEOUS AND KINESTHETIC SENSATIONS	933
CHAPTER 64	
REFLEX ACTION	948
CHAPTER 65	
THE PHYSIOLOGICAL MECHANISMS GOVERNING POSTURE AND EQUILIBRIUM	960
CHAPTER 66	
THE SPINAL CORD AND BRAIN STEM (MEDULLA, PONS AND MID-BRAIN)	986
CHAPTER 67	
THE EXTRAPYRAMIDAL SYSTEM THE THALAMUS AND HYPOTHALAMUS	1015

CHAPTER 68

THE CEREBRAL CORTEX. THE PHYSIOLOGY OF SPEECH AND SOME OF ITS DISORDERS APRAXIA AND AGNOSIA EPILEPSY HEADACHE	1030
--	------

CHAPTER 69

CONDITIONED REFLEXES SLEEP	1057
----------------------------	------

CHAPTER 70

THE CEREBELLUM	1072
----------------	------

CHAPTER 71

THE CEREBROSPINAL FLUID (C S P)	1084
----------------------------------	------

CHAPTER 72

THE AUTONOMIC NERVOUS SYSTEM	1088
------------------------------	------

SECTION IX THE SPECIAL SENSES

CHAPTER 73

THE PHYSIOLOGY OF VISION STRUCTURE OF THE EYEBALL DUPLICITY THEORY REACTIONS OF THE RETINA, SUBJECTIVE PHENOMENA	1106
---	------

CHAPTER 74

THE NATURE OF THE RETINAL PROCESS, OBJECTIVE RETINAL PHENOMENA, COLOR VISION, OPTICAL ILLUSIONS	1127
--	------

CHAPTER 75

THE DIOPTRIC MECHANISMS OF THE EYE CATARACT OPTICAL DEFECTS INTRA- OCULAR FLUIDS	1141
---	------

CHAPTER 76

THE VISUAL FIELDS AND PATHWAY THE PUPIL MOVEMENTS OF THE OCULAR MUS- CLES STEREOSCOPIC VISION	1167
--	------

CHAPTER 77

THE EAR ANATOMICAL OUTLINE SOUND GENERAL PRINCIPLES	1185
---	------

CHAPTER 78

AUDITORY SENSATIONS THE MECHANISM OF HEARING THE TELEPHONE AND RESO- NANCE THEORIES	1195
--	------

CHAPTER 79

THE EUSTACHIAN TUBE DEAFNESS HEARING TESTS THE HARMFUL EFFECTS OF NOISE	1209
--	------

CHAPTER 80

THE CHEMICAL SENSES	1213
---------------------	------

REFERENCES	1225
------------	------

INDEX	1295
-------	------

SECTION I. THE BLOOD AND LYMPH

By N B T

CHAPTER 1

THE PHYSIOLOGICAL PROPERTIES, PHYSICAL CHARACTERS AND COMPOSITION OF THE BLOOD

OUTLINE OF THE FUNCTIONS OF BLOOD

In animals whose bodies are composed of many cells (Metazoa) the blood serves those purposes, which for unicellular organisms (Protozoa) are carried out by the fluid medium, the salt or fresh water, which surrounds them and bathes their surfaces. For example, an organism such as the amoeba acquires oxygen by diffusion directly from the environment into the interior of the cell. Similarly the carbon dioxide diffuses outwards. The processes of nutrition and the excretion of the products of the cell's metabolism are accomplished in a manner equally simple. Food is taken in through the cell membrane either in solution or as particulate matter, and waste products pass into the surrounding medium. Other requirements of this organism, such as the maintenance of an optimum temperature and the proper degree of moisture, are dependent on the immediate (internal) environment, the *milieu interne* of Claude Bernard.

The elemental needs of each cell in a multicellular form from the most primitive type to the highest vertebrate are the same as for the unicellular organism, yet in the evolution of the higher forms the cells composing their bodies have become farther and farther removed from immediate contact with the outside world. Myriads of cells have become packed together, and the deeper ones could not possibly satisfy their needs after the direct and simple fashion of the unicellular forms. The more primitive multicellular types overcame the difficulty by the development of canal systems which opened to the exterior and through which the ocean waters flowed freely in and out, bringing oxygen and aliment to the more deeply lying cells and bearing carbon dioxide and other excretory products away. This, the first attempt at a circulation, was an open one. As higher forms evolved the circulation became closed and the waters of the environment no longer flowed and ebbed through the body. No longer could the interchange of the respiratory gases and the absorption of nutriment be carried out in this direct

and simple way. Yet the vessels of this closed circulatory system were filled with a fluid which took the place of and fulfilled the duties of the watery environment of the more primitive types. The blood and other body fluids may be looked upon as that environment which has become enclosed within the bodies of the higher forms, and has undergone certain modifications in its composition to meet the requirements of the more specialized cells which it bathes.

The similarity between the compositions of sea water and blood which has been stressed by the researches of Macallum lends support to these views on the evolution of the blood.¹ This brief account will also serve as an introduction to a consideration of the functions of the body fluids, since their duties are to satisfy in the same way as did their prototype, the requirements of the individual cells.

(1) *Respiratory* The transport of oxygen from the air in the lungs to the tissues, and of carbon dioxide from the tissues to the lungs.

(2) *Nutritive* The conveyance of food materials, glucose, amino acids and fats from the alimentary canal to the tissues.

(3) *Excretory* The removal of waste products of metabolism, e.g., urea, uric acid, creatinine, etc.

(4) *The maintenance of the water content of the tissues* Though the blood itself is contained within vascular channels, a constant interchange of fluid through the vessel walls takes place. This fluid which has left the blood vessels and come into direct contact with the tissue cells is known as the tissue or interstitial fluid. It closely resembles the blood plasma in chemical composition, and is identical with lymph. Through the medium of the transuded fluid the final stage in the transportation of oxygen and food materials to the tissues and the

¹ Sea water of today differs from blood serum in having a total salt concentration of about 3 per cent, a much higher concentration of magnesium and a lower concentration of potassium. But Macallum points out that the sea water of the geological period when the ancestors of mammalian forms adapted themselves to a terrestrial life was probably closely similar in its inorganic composition to blood serum.

first stage in the journey of CO_2 and waste products from the tissues are made

(5) *To regulate body temperature* The body owes its ability to regulate its temperature (ch. 53) largely to the water of the blood and tissue fluids. Water possesses three qualities which fit it pre-eminently to fulfil this purpose.

(a) The *specific heat*^{*} of water is considerably higher than that of any other liquid or solid. On account of this great heat storage power of water, sudden changes of body temperature are avoided and even a cold-blooded animal such as the frog has, due to this purely physical quality, some ability to maintain a relatively constant body temperature against transient fluctuations in environmental temperature. A man of average weight develops 3600 Calories in 24 hours. This amount of heat is capable of raising the temperature of his tissues (which are largely water) only about 32°C . Heat elimination (radiation, etc.) is able to keep pace with heat production and the body temperature varies but slightly within normal limits. But it has been pointed out by L. J. Henderson that if the tissues had the low heat storage capacity (spec. heat) of most substances, an amount of heat equal to 3600 Calories would raise the temperature of the tissues and fluids of the body by from 100° – 150°C .

(b) *High conductivity* The thermal conductivity of water is greater than that of any other ordinary liquid. The advantage of this in the dissipation of heat from deeply situated regions of the body is obvious.

(c) *High latent heat of evaporation* More heat is required for the vaporization of water than for that of an equivalent amount of any other liquid. One cc of water requires about 0.6 Calories for its vaporization. This figure is 50 per cent higher than that of water's closest competitor. Fluid is being constantly lost from the body through evaporation from the lungs and skin. A large amount of heat is lost in the process (ch. 53).

These physical properties of water which make it ideal as a heat regulating medium are enhanced by other purely *physiological factors*. The mobility of the blood, and the readiness with which it may be quickly redistributed in the body, combined with the unique physical properties of the fluid itself, render it so highly efficient as a regulator of body temperature. The blood may in a moment be brought from deeper to superficial regions and spread out in fine vessels over a broad area just beneath the skin, and in this way will greatly increase the radiation of heat. At another instant, in order that heat may be conserved, the fluid is drained from the surface areas and collected in the deeper parts of the body—internal organs, muscles, etc.

(6) *Protective and regulatory* The blood and lymph contain certain chemical substances of a complex nature, antitoxins, lysins, and other antibodies, which

are the basis of the body's defence against injurious agents of various kinds. The circulating fluids are also the vehicle by which the hormones of the different ductless glands are brought into direct contact with the cells of the tissues.

THE COMPOSITION OF BLOOD

The blood is a highly complex fluid in which solid elements are suspended—the *corpuscles* or *blood cells*. Its specific gravity is from 1.050 to 1.060 and its viscosity from 5 to 6 times that of water. If blood is centrifuged before it has had time to clot, or if clotting is prevented by special means (p. 117), the solid elements are thrown down and separated from the fluid portion. The latter is called the *plasma* and contains *proteins*, as well as many organic and inorganic substances in solution—nutritive and excretory materials, antibodies and hormones, and other substances of an unknown or imperfectly known chemical constitution. The specific gravity of plasma is normally around 1.027 but varies with its concentration in protein. The cells constitute about 46 per cent of the volume of human blood, the plasma 54 per cent. Small variations above or below these values are commonly met with.

The specific gravity of a small sample of blood or of plasma may be measured by the method of Phillips, Van Slyke and associates. A series of small bottles is set up containing copper sulfate solutions varying by small equal increments (0.004) in specific gravity. A drop of blood or plasma is allowed to fall gently from the tip of a medicine dropper into each of a number of the bottles whose solutions are within the expected specific gravity range of the blood or plasma sample. The tip of the medicine dropper should be held about 1 cm above the surface of the solution. The drop of blood or plasma, upon entering the solution, becomes coated with a film of copper proteinate and remains suspended, neither rising nor falling for a few seconds, if it is of the same specific gravity as the solution, thus, since the specific gravity of the solution is known, that of the blood or plasma is indicated (fig. 11).

In the following table are given the constituents of the blood, grouped upon a physiological basis.

Whole blood

A. Cells

- (1) Red corpuscles or erythrocytes
- (2) White corpuscles or leucocytes
- (3) Platelets or thrombocytes

* The specific heat of a substance is defined as the number of calories required to raise 1 gram of the substance one degree Centigrade.

B Plasma

- (1) *Water*, 91 to 92 per cent
- (2) *Solids*, 7 to 9 per cent
 - (a) *Proteins*, 7 per cent Serum albumin, serum globulin and fibrinogen²
 - (b) *Inorganic constituents*, 0.9 per cent Sodium, calcium, potassium, magnesium, phosphorus, iodine, iron, copper, etc.
 - (c) *Organic constituents* (other than (a) and (d)) Non-protein nitrogenous substances, (urea, uric acid, xanthine, hypoxanthine, creatine and creatinine, ammonia and amino acids) neutral fats, phospholipids, cholesterol, glucose
 - (d) *Internal secretions, antibodies and various enzymes*, (amylases, proteases, lipases, esterases, etc.)

INORGANIC CONSTITUENTS

The concentration of the plasma in the various inorganic materials is given in table 1

It will be noted that the plasma is relatively rich in sodium and calcium but poor in potassium and magnesium whereas in the cells conditions are reversed. The cells show a relatively high concentration in potassium and magnesium, but are lacking in calcium and have a low concentration of sodium (human). In the blood of some species sodium is absent or present only in traces. Except for a minute amount of iron in the plasma, this element is confined to the red cells and the greater part of it is attached to the hemoglobin molecule (ch 6). It has been suggested that the small quantity of non-hemoglobin iron in the erythrocyte is bound loosely with the lecithin of the cell stroma (p 9) (See also ch 8)

Phosphorus

Phosphorus exists in blood in four main forms. One of these is *inorganic phosphorus* (orthophosphate). The three other phosphorus fractions are in *organic* combination and are as follows:

- a *Ester phosphorus*, e.g., diphosphoglycerate, adenosinetriphosphate, hexose phosphates, glycerophosphate
- b *Lipid phosphorus*, e.g., the phosphatides lecithin, cephalin, sphingomyelin
- c *Nucleic acid phosphorus*

According to Kay the nucleic acid phosphorus in normal human blood is negligible. It is derived from the nuclei of white cells and the reticulum of the reticu-

² Plasma from which the fibrinogen has been removed through clotting (ch 12) is spoken of as serum



FIG 11 Description in text (After Phillips and Van Slyke, redrawn)

locytes. In abnormal blood containing a large number of leucocytes, reticulocytes or nucleated red cells this fraction may however constitute a considerable proportion of the total phosphorus.

The inorganic phosphorus (3 mg per 100 cc) is according to most observers about equally distributed between cells and plasma. The quantity of organic phosphorus in blood is many times greater than the inorganic. In whole blood it amounts to from 35 to 40 mg per 100 cc and the greater proportion of this is in the cells.

The inorganic and ester fractions are extracted from blood by the precipitation of the proteins with trichloroacetic acid and filtering. The phosphorus contained in the filtrate is spoken of as the acid soluble phosphorus. Upon extraction of blood with alcohol-ether the lipid phosphorus is obtained. The phosphorus of blood is therefore separable into two classes:

- (1) *The acid soluble which includes*
 - (a) Inorganic phosphorus
 - (b) Ester phosphorus
- (2) *Alcohol-ether soluble, i.e.,* } organic phosphorus
lipid phosphorus

TABLE 1

Inorganic constituents of plasma, red cells and whole blood, milligrams per 100 cc average values

	SODIUM	POTAS- SIUM	CAL- CIUM	MAGNE- SIUM	CHLO- RINE	IODINE	IRON	COPPER	PHOS- PHATE	SUL- FATE	TOTAL BASE CC N/10 NaOH
Plasma	340	20	10	2 7	370		0 2				160
Cells	20	410	0	6 0	190		100 0				
Whole blood	190	220	5 2	4 0	250	0 01	50 0	0 1	3 0	2 0	

The concentrations of these various inorganic constituents are also commonly expressed as milli-equivalents (m eq) per liter. Thus serum contains 100 mg of calcium per liter. The molecular weight of Ca is 40.07, being divalent its milli equivalent is 20.03. The concentration of calcium in serum is therefore $\frac{100}{20.03} = 4.9$ mil equivalents per liter. Sodium is monovalent and has a molecular weight of 23, serum therefore contains $\frac{3400}{23} = 147.8$ m eq per liter.

The ester, or organic acid soluble phosphorus is obtained by determining the total acid soluble P and subtracting from it the inorganic phosphorus. Of the ester phosphorus, all of which is intracellular, about one-quarter is hydrolyzable by bone phosphatase (ch 60). The hydrolyzable portion is mainly adenosinetriphosphate and the non hydrolyzable part mainly diphosphoglycerate. Since the nucleic acid phosphorus is negligible in normal blood, the acid soluble + the alcohol ether soluble phosphorus equals the total phosphorus as determined by wet ashing.

In the following table is given the distribution of inorganic, ester and lipid phosphorus in normal blood.

Phosphorus in whole blood

Milligrams per 100 cc, average figures

1 Total phosphorus	40
2 Total acid soluble—90 per cent in cells	27
3 Inorganic—in cells and plasma	34
4 Ester (2-3)—practically all in cells	24
5 Lipid (1-2)—in cells and plasma	13

The phosphorus compounds of the blood and tissues play an important role in maintaining the electrolyte equilibrium within the red cells and in regulating the acid base balance. Diabetic acidosis, for example, and the acidosis induced by the ingestion of ammonium chloride, are accompanied by increased excretion of phosphorus in the urine and a pronounced reduction of the organic acid-soluble phosphorus in the blood cells. Reverse changes occur in alkalosis, the reduction in the chloride of the blood following pyloric obstruction, and the alkalosis caused by over-breathing are associated with a reduction in the urinary excretion of phosphates and a decrease in the inorganic and ester phosphorus.

⁴ In infants and young children, the inorganic phosphorus is from 1 to 3 mg per cent higher than it is in adults.

of the blood. In renal insufficiency, the inorganic phosphorus in the plasma and cells and the ester phosphorus (diphosphoglycerate) in the cells are greatly increased. The inorganic and ester phosphorus are reduced in rickets but a rapid increase accompanies the healing process. The inorganic phosphorus is diminished after the injection of insulin and in hyperparathyroidism (ch 60). In anemias associated with high reticulocyte counts and in leukemia, the concentration of ester phosphorus in the blood is increased. The inorganic phosphorus is increased in some forms of tetany.

ORGANIC CONSTITUENTS (OTHER THAN
ORGANIC PHOSPHORUS)*Plasma proteins*

The concentration of total protein in the plasma and the proportions of the three fractions—albumin, globulin and fibrinogen—vary from species to species but under ordinary conditions of health remain relatively constant between individuals of the same species.

Serum globulin can be separated by "salting out" into two fractions—euglobulin and pseudoglobulin, or into three fractions— α , β and γ -globulins—by electrophoresis. The euglobulin is thrown out of solution by saturation with NaCl, half-saturation with $MgSO_4$, or one-third saturation with $(NH_4)_2SO_4$, it is insoluble in water. The pseudoglobulin is not "salted out" by NaCl but thrown down by saturation of its solution with $MgSO_4$ or half-saturation with $(NH_4)_2SO_4$. It is soluble in water.

α , β and γ -globulins have isoelectric points 5.1, 5.6 and 6.0 pH, respectively. It is questionable whether these fractions are distinct chemical entities. It is more probable that they are merely

artificially produced as a result of the methods of treatment employed. In other words, it is likely that serum globulin is a single large molecule, which is split into two or three separate parts by laboratory manipulation. Yet however this may be, the gamma-globulin is more intimately associated with antibody production, and undergoes an increase in many acute and chronic infections.

Pseudoglobulin contains 85 per cent alpha-globulin and 15 per cent gamma-globulin, whereas, euglobulin contains less alpha but more of the beta and gamma globulins.

The several electrophoretic⁶ fractions of plasma protein are not pure, all contain lipid and carbohydrate material combined probably as prosthetic groups. The albumin fraction also contains bilirubin (ch 40). It has been estimated that at least 50 per cent of the lipid and carbohydrate content of serum is bound to the albumin and gamma-globulin fractions. Other substances, e.g. calcium, phosphorus, sulfonamide drugs and the dye T-1824 (p 18) are bound to the albumin fraction.

Fibrinogen has been isolated and prepared in crystalline form. X-ray diffraction studies indicate that its molecule is structurally similar to such fibrous proteins as collagen and myosin (see p 622). The molecular weights of the plasma proteins are given in chapter 35.

The total plasma protein can be calculated from the specific gravity of the plasma by means of line charts, or by using the formula $P = K(S - A)$, where P is the plasma protein in grams per 100 cc, S the specific gravity and K and A are constants with values of 364 and 1.006, respectively. Thus, if the specific gravity is 1.026, the protein in grams per 100 cc is 7.28 ($364(1.026 - 1.006)$).

The values of total protein and of the different fractions in human plasma are given in the following table.

Protein fractions in human plasma

FRACTIONATION BY ELECTROPHORESIS		FRACTIONATION BY SALTING OUT WITH SODIUM SULFATE	
	grams/100 cc		grams/100 cc
Total protein	6.03-6.72	Total protein	6.0-8.0
Albumin	3.32-4.04	Albumin	4.3-5.0
Total globulin	2.23-2.39	Total globulin	1.1-3.1
Alpha globulin	0.79-0.84	Euglobulin	0.1-0.4
Beta globulin	0.78-0.81	Pseudoglobulin	1.0-2.7
Gamma globulin	0.66-0.70		
Fibrinogen	0.34-0.43	Fibrinogen	0.2-0.3
		Albumin/globulin (A/G) ratio	1.50

In some animals the globulin is equal to or exceeds the albumin. Of the three fractions fibrinogen is always in lowest concentration and it is considerably lower in human plasma than in that of some animals (e.g., 0.58, 0.72, 0.60 gram per 100 cc in dog, cow and goat respectively).

PATHOLOGICAL VARIATIONS IN CONCENTRATION

The several protein fractions of plasma may change in value independently of one another, and either with or without alteration in the quantity of total protein, in several pathological states the albumin and globulin fractions may change in opposite directions, i.e. a fall in albumin accompanied by a rise in globulin.

The fibrinogen concentration is increased in pregnancy and menstruation, in tissue injury of various kinds, in parathyroid overdosage, acute infections, malaria and several other conditions. This fraction is markedly reduced in animals after hepatectomy or severe liver damage and in several diseases involving the liver. In rare instances it is congenitally considerably below normal or absent.

In hemorrhage a loss of all fractions of plasma protein occurs, their concentrations are also diminished as well, since the blood volume is at first made good by the passage of a saline solution or one of low protein concentration from the tissue spaces into the blood stream. In extensive burns, on the other hand, especially during the following few days, all fractions are reduced as a result of the leakage of blood fluid from the denuded surface and into the tissues in the region of the burned area, but since the lost fluid is usually relatively low in its content of protein, the protein concentration of the plasma tends towards an increase. In cirrhosis of the liver, chronic hepatitis (depressed

⁶ The usual method used today in electrophoresis (i.e. the migration of charged particles in an electric field to cathode or anode) of protein in a suitable buffer solution and other colloidal systems is that carried out with the apparatus of Arn Tiselius, in which the moving boundaries formed between the protein and buffer solutions are recorded graphically by optical methods. The different proteins of serum are separable upon the basis of the rates and direction of movement of their boundaries. The most rapidly moving boundary is that of the smaller albumen molecule. Alpha, beta and gamma-globulins have much slower rates but of the three the alpha fraction has the fastest rate, and the gamma globulin the slowest. By ultra-violet photography a characteristic electrophoretic pattern of these boundaries is obtained, which shows a series of peaks corresponding to the individual proteins in the solution. The albumin peak is by far the highest.

synthesis by liver), *chronic infections*, the albumin fraction is reduced

In *nephrotic* and *nephritic conditions* (due to loss of albumin in the urine) and in severe *malnutrition* (owing to the low intake of the necessary amino-acids for protein synthesis) the albumin fraction is also reduced. As a result of the diminished concentration of albumin, the oncotic pressure of the plasma tends to fall, less water is held in the vessels and, as a consequence, the plasma volume is reduced. As a consequence of these changes, the globulin, though not raised absolutely, shows increased concentration. In any condition associated with a loss of water from the blood (*dehydration*, *anhidremia*, p 24), though no change may occur in the absolute amount, i.e. in the total quantity in the blood, the concentrations of all fractions will show an increase. In order to determine whether or not an *absolute* reduction or increase in one or other of the plasma proteins exists, it would be necessary to measure the total plasma volume (p 18) as well as the concentration of the particular fraction. The gamma-globulin shows an absolute increase in *multiple myeloma*, *cirrhosis of the liver*, *subacute yellow atrophy of the liver*, *acute hepatitis* and *acute nephritis*, *leukemia*, *tuberculosis*, *scarlet fever* and in *acute* and *chronic infections*. In liver disease the alpha-globulin is also increased, and in later pregnancy the concentration of beta-globulin, which contains a high percentage of lipid, is raised. Prothrombin appears to be a beta-globulin. This fraction serves to bind a considerable proportion of the plasma cholesterol, carotene and phospholipids of the plasma, and is increased in the later months of pregnancy. The isoagglutinins (anti-A and anti-B, and anti-Rh, ch 5) are associated with the gamma and beta globulin fractions of the plasma.

ORIGIN In the *embryo*, the mesenchyme cells through a process of secretion or by the actual solution of their substance furnish the fluid (embryonic plasma) which floats the primitive blood cells (p 103). The albumin fraction is formed earlier than the other proteins which do not appear in the plasma of the chick embryo until after the 14th day of incubation.

In the *adult*, five possible sources of the plasma proteins have been suggested—namely, disintegrating blood cells (red or white), the general tissue cells, reticuloendothelial cells of spleen, bone marrow, etc., and the liver.

It is now well established that the liver is the site of the production of plasma albumin, this fraction

undergoes a pronounced reduction in conditions which depress hepatic function. It is thought that the Kupffer cells are especially concerned in the manufacture. The albumin fraction (but not the globulin) in the plasma of dogs can be reduced and maintained at a subnormal level by intravenous injections of a solution of gum acacia (p 50). In these experiments the hepatic cells become swollen and vacuolated. The fall in serum albumin is therefore considered to be due, in part at least, to the failure of the liver to make good the normal "wear and tear" of serum albumin (which amounts to several grams daily), though, possibly, there is also a withdrawal of albumin from the plasma to the liver as a compensatory response to rectify the increase in oncotic pressure caused by the presence of acacia in the circulation.

The evidence points definitely to the liver as the site of fibrinogen production since as mentioned above the concentration of this fraction is reduced by liver damage or hepatectomy. When the liver returns to a healthy state after injury has been induced by an agent such as phosphorus or chloroform, the fibrinogen level also returns to normal. Following slight liver injury or during the repair of a hepatic lesion, which might be expected to stimulate the functional activity of the organ, the fibrinogen may be actually higher than normal.

The origin of serum globulin has not been definitely settled. Elman and Heifetz succeeded in reducing the albumin fraction in the serum of dogs by dietary measures to 50 per cent of the normal. No reduction was observed in the serum globulin. The livers of the animals showed pronounced histological and chemical changes, the water content of the hepatic tissue was increased, the cells became vacuolated. This was accompanied by severe depletion of hepatic protein.

On the contrary, the experiments of Miller and his associates who perfused the intact livers of rats with blood containing lysine labeled with isotopic carbon (C^{14}) indicate that the liver furnished about 80 per cent of the serum globulin. The remainder is supplied by extrahepatic tissues, the greater part of which, if not the whole, are derived from the lymphocytes (White and Dogherty, p 96). The conflict between the results of the experiments of Elman and Heifetz and those of Miller and associates may be explained perhaps by the great functional reserve of the liver, hepatic lesions in the experiments of the former being insufficient to cripple its globulin producing function.

There is experimental evidence that the adreno-

corticotrophic hormone of the pituitary through its action on the adrenal cortex affects the manufacture of beta and gamma globulins. Injections of this hormone or of adrenal cortical hormones into rats causes within 24 hours a rise in beta globulin of 30 per cent, and in gamma globulin of from 70 to 80 per cent (White and Dougherty)

FUNCTIONS (1) Fibrinogen is essential for the clotting of the blood (ch 12)

(2) All three proteins serve to maintain the osmotic pressure (p 29) of the blood. The large molecules of the proteins do not pass readily through the normal capillary membrane. The osmotic pressure which they exert amounts to, in man, between 25 and 30 mm Hg. The pressure which each fraction exerts is inversely related to the size of its molecule and directly related to its concentration in the plasma. The molecular weight of fibrinogen is over 200,000 and its concentration is low, it therefore contributes little toward the total osmotic pressure. Albumin is in the highest concentration and its molecule has the least weight (70,000-75,000). The osmotic pressure of the plasma, therefore, depends largely upon this fraction. The molecular weight of serum globulin is between 150,000 and 190,000 and its concentration is considerably less than that of albumin. In equivalent concentrations, serum albumin has an osmotic activity 2.4 times that of serum globulin (Keys), it furnishes about 80 per cent of the oncotic pressure of the plasma.

(3) Viscosity. The proteins give a certain viscosity to the blood which is a factor in the maintenance of the normal blood pressure (p 149)

(4) They aid in the regulation of the acid-base balance of the blood (ch 13)

(5) Stability of the blood (see p 66). The globulin and fibrinogen fractions influence the tendency of the corpuscles to adhere to one another and form rouleaux or clumps.

(6) Trephones. Carrel has shown that the leucocytes prepare substances from the plasma proteins which are essential for the nourishment of tissue cells grown in cultures. These substances he has termed trephones.

(7) Immune substances (antibodies) which react with the antigens of several microorganisms, e.g. diphtheria, typhoid and streptococcal infections, and the viruses of mumps, influenza and measles, are associated with the gamma globulin. Gamma globulin, separated from the other fractions, is used as a means of artificially immunizing against measles, infectious hepatitis, rubella and polio-

myelitis.⁶ Certain other antibodies, the isoagglutinins A and B, are present in the gamma and beta fractions. As might be expected, the globulin fraction of the serum tends to increase during the process of immunization against the infective diseases mentioned above.

(8) They serve as a reserve of protein upon which for a time the body draws during fasting or when the protein intake is inadequate.

Plasmapheresis. The importance of the plasma proteins is demonstrated by this procedure which consists in bleeding an animal and returning the red cells suspended in Locke's solution to the body. A state of shock results, followed by death when the total protein is reduced to between 1 and 2 per cent. No ill effects result however if the cells are suspended in serum before they are reintroduced. When depletion of the proteins is not carried to the point where fatal shock ensues, a marked rise in protein concentration occurs within 15 minutes which indicates that during this time a store of pre-formed protein is drawn upon for the replacement of the protein which has been removed. The regeneration is slower after this, though fairly rapid for the first 24 hours. It becomes progressively slower during succeeding days. The proteins are restored to the normal level in from 2 to 7 days, provided that the diet contains a sufficiency of high quality protein. Plasma proteins themselves have been found to be best for this purpose, the proteins of liver run a close second.

In more chronic plasmapheresis experiments, edema commences when the total protein concentration reaches a value of 5.5 per cent and albumin a concentration of 2.5 per cent.

The non-protein nitrogen (N.P.N.) or non-coagulable nitrogen of blood

By the term non-protein-nitrogen is meant the nitrogen of those substances e.g. urea, uric acid, creatinine, etc., listed on page 8. They may be extracted from blood or plasma by treating either of these with a reagent, such as trichloroacetic acid, which precipitates the proteins, filtering and determining the nitrogen in the filtrate. These substances are in part absorbed with, or derived from the food, and in part are the waste products of tissue catabolism. The total N.P.N. of whole blood amounts to from 28 to 40 mg per 100 cc. It

⁶ Gamma globulin obtained from any normal person usually contains sufficient immune bodies for protection against measles and infectious hepatitis, but for protection against rubella (German measles), poliomyelitis, whooping cough, chickenpox and mumps, an active immune globulin can be obtained only from subjects convalescent from the particular disease.

TABLE 2*

The nitrogen partition in the blood of normal individuals and the distribution of the various nitrogenous constituents between the cells and serum

	CORPUSCLES			PLASMA			WHOLE BLOOD		
	Maxi- mum	Mini- mum	Aver- age	Maxi- mum	Mini- mum	Aver- age	Maxi- mum	Mini- mum	Aver- age
	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc
(a) Taken from Wu									
Total non protein nitrogen	61	39	49	36	20	29			
Urea N	22	12	17	23	13	19			
Amino-acid N	11	8	10	8	5	6			
Uric acid	4	1	2	5	2	4			
Creatine	8	4	6	0	0	0			
Creatinine	3	1	2	5	1	1			
Undetermined N			19			2			
(b) Taken from Berglund									
Total non protein nitrogen	55	38	44	30	18	25	39	28	32
Urea N	13	8	10	17	10	12	15	9	12
Amino-acid	11	7	8	6	4	5	8	6	6
Undetermined N	34	18	25	12	2	7	18	10	14

* Reprinted from Peters and Van Slyke, *Quantitative Clinical Chemistry*, vol I, p 267 The Williams & Wilkins Company, Baltimore, 1946

constitutes from 1 to 2 per cent of the total nitrogen of the blood. Its concentration in the cells is nearly double that in the plasma. The proportions of the different constituents are given in table 2.

The term undetermined nitrogen embraces the nitrogen of ammonia, purines, and other non-protein substances of unknown or undefined nature.

PATHOLOGICAL VARIATIONS The N P N of blood at any given level represents the balance struck between nitrogenous materials formed in the intermediary metabolism of ingested and tissue protein and the excretion of these products in the urine. In renal insufficiency, therefore, the non-protein nitrogen is elevated, and in certain cases may be ten times the normal. On the other hand, a rise in the N P N occurs in conditions which are associated with excessive tissue catabolism, such as infections, fevers, thyrotoxicosis, starvation or severe malnutrition. It is also increased following hemorrhage into the stomach or upper intestinal tract. In the later months of pregnancy the N P N is reduced. The reduction has been attributed to the diversion of nitrogen to the growing fetus and

the reduction of protein catabolism in the maternal tissues.

The chief conditions associated with an elevation of the N P N of the blood are:

- Adrenal insufficiency
- Dehydration
- Hemorrhage into the gastrointestinal tract
- Infectious fevers, lobar pneumonia
- Intestinal obstruction
- Parathyroid intoxication (in animals)
- Peritonitis
- Renal insufficiency

Cholesterol

(See also chaps 50 and 55) This sterol is present in serum in the free form (45 mg per cent) and as cholesterol esters (110 mg per cent). Its concentration is increased in several diseases, notably in atheromatosis, hypothyroidism and lipid nephrosis. Keys and his associates found that within fairly wide limits (2 to 3 grams cholesterol weekly) the content of the sterol in the diet exerted little effect upon the serum cholesterol, though hypercholesterolemia can be considerably reduced by a cholesterol free diet.

CHAPTER 2

THE RED CELLS OR ERYTHROCYTES

THE SIZE, SHAPE AND STRUCTURE OF THE RED CELL

Human erythrocytes are disc-shaped, non-nucleated elements having a mean diameter of 7.2 microns (6–9 μ , see fig. 2.1) and a thickness of about 2.2 (2–2.4 μ) microns at the thickest part, i.e. near the circumference, and about one micron at the center. As a result of osmotic changes¹ and the consequent passage of water into the cell, the diameter increases with a shift in the acid-base balance of the blood toward the acid side. The cell is therefore slightly larger in venous than in arterial blood, its diameter is increased by about 0.5 μ in muscular exercise and reduced by forced breathing. The central portion of the cell is much thinner than its edges, which appear heaped up into a circumferential mound around a central depression. This construction gives it a biconcave contour or a roughly dumb-bell outline when viewed edgewise (figs. 2.1 and 2.2).² The average area of a red cell is 120 square microns and the volume 85 cu. microns. The mature erythrocyte can scarcely be considered a living cell in the ordinary sense, since it possesses no nucleus and does not consume an appreciable

¹These are the mean dimensions of the cell measured in dry films. In the fresh state the diameter is larger by about 0.5 microns. There is considerable variation between the diameter of the smallest and largest cells found in a sample of normal blood. The range for dried films is shown in figure 2.2.

²Hartridge has pointed out the advantage of this design for the transport of oxygen. Of all geometrical figures the sphere is the one in which its centre is equidistant from all points upon its surface. The adoption of this form by the red cell would therefore have ensured the diffusion of oxygen to all parts of its interior at equal rates. But a sphere has the disadvantage of possessing the smallest surface in relation to its mass. A thin disc, on the other hand, presents an almost maximal surface area in relation to its bulk, yet in such a shape all parts on the surface are not equally distant from its center, the ends are further removed than the sides. The shape of the red cell—a thin disc with elevated rounded edges—is a compromise between these two forms. It secures the advantages of equal and rapid diffusion of oxygen to its interior and a relatively large surface area for the absorption of the gas.

The biconcave form also gives the red cell a mechanical advantage, in that the changes in volume which the cell undergoes from time to time can be effected with a minimal amount of tension being placed upon the cell membrane. The membrane covering the concavity of the cell moves freely out or in "like the bottom of an oil-can" as the cell increases or diminishes in volume.

amount of oxygen. Young (nucleated or reticulated) cells, on the other hand, and the nucleated cells of lower vertebrates consume a considerable quantity of oxygen.

The red cell is bounded by a membrane made of protein in association with lipid and steroid materials, chiefly lecithin and cholesterol. Though this membrane has not been demonstrated histologically, indirect evidence indicates its presence and suggests that it consists of an outer and inner layer of protein, each probably only a few molecules thick and enclosing between them three or four layers of lipid. It behaves as a semipermeable membrane. The body of the cell possesses a sponge-like stroma made of the same or similar materials probably in the form of a gel. The protein is a paraglobulin, small quantities of nucleoprotein are also

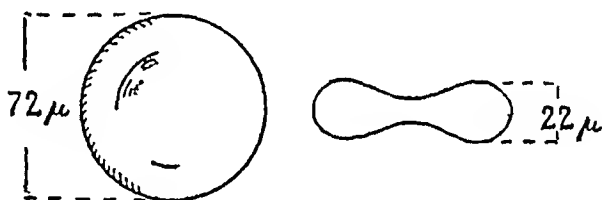


FIG. 2.1 Diagram showing dimensions of the red cell

present. In the meshes of the stroma, or more likely actually bound up in the stroma substance itself, the respiratory pigment *hemoglobin* is held. The fact that mechanical division of the cell, even into the finest particles, fails to liberate the hemoglobin, supports the conception of a very intimate association of the pigment with the cell stroma. The water content of the cell is lower than that of most cells of the fixed tissues, amounting to about 60 per cent (53–63 per cent). Hemoglobin makes up from 80 to 90 per cent of the total solids of the cell (and about 34 per cent of its fresh weight). Other proteins (0.5 to 1 per cent), phospholipids (lecithin and cephalin) and cholesterol (0.4 and 0.3 per cent respectively), inorganic salts (fig. 2.3), urea, amino acids, creatine, etc. make up the remainder of the cell solids. Potassium is the principal base in the human erythrocyte, and sodium in the red cells of the cat and dog. The specific gravity of the red cell is 1.091.

When a drop of freshly drawn normal blood is

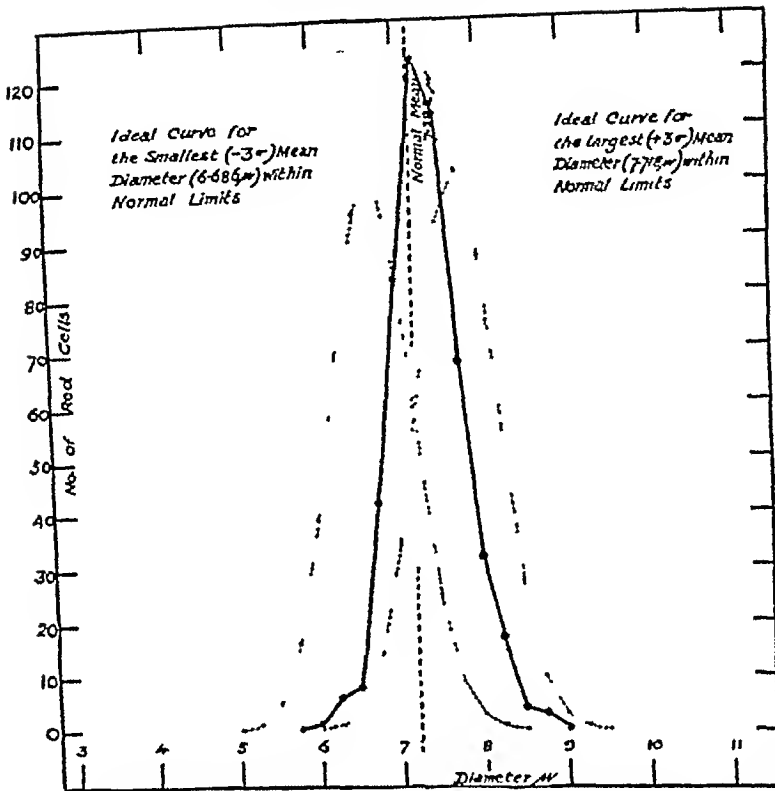


FIG 2.2 Red-cell-diameter distribution curve for healthy men (after Price-Jones)

placed on a glass slide and examined under the microscope, many of the erythrocytes will be seen to group themselves together with their broad surfaces in contact, like a pile of coins. Groups of cells arranged in this way are called *rouleaux* (sing *rouleau*). The normal discoid shape of the erythrocytes is a requisite for *rouleaux* formation, the property being lost if the cells, as in congenital hemolytic jaundice, assume a more globular form. *Rouleaux* formation does not occur in the circulation under normal physiological conditions, the moving cells showing little or no tendency to cohere.

"Sludged" blood In certain abnormal states, e.g., tissue injury, and shock, the cells of the circulating blood show a pronounced tendency to stick together and form large clumps or masses which move slowly and clumsily through the small vessels—arterioles and capillaries. This "sludged" blood, as it is termed by Knisely who has made an extensive study of the phenomenon, has a deleterious effect upon the general circulation. Normally the blood moves through the minute vessels in stream lines. The cells remain discrete showing little tendency to cohere. The

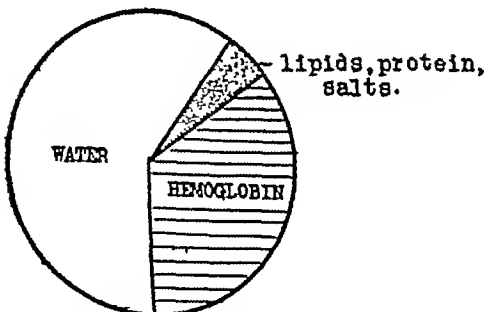


FIG 2.3 Diagram showing composition of the red cell.

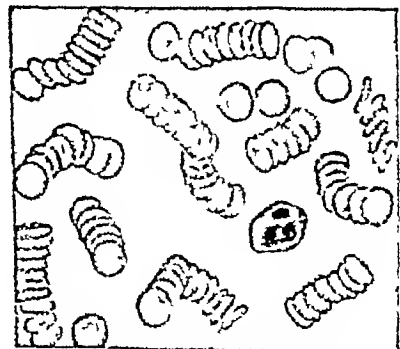


FIG 2.4 Description in text.

propulsive force of cardiac systole is transmitted throughout the vascular system with the minimum dissipation of energy. Stream-lining is abolished in sludged blood, a large proportion of the energy is wasted in giving a rolling or rotary motion (angular acceleration) to the massed cells. Sludged blood is thought in some instances to be a factor in thrombosis (ch 12).

NUMBER

The average number of red cells in man is around 5,000,000 per cubic millimeter for males and 4,500,000 for females, but 6,000,000 is not a very unusual figure for a robust young man and 5,500,000 for a healthy young woman.³ Slight variations in the number of red cells, amounting to about 5 per cent, occur throughout the twenty-four hours. The count is lowest during sleep, becomes elevated after arising and increases gradually throughout the waking hours. At birth and in infancy the red cells are somewhat more numerous than in later life, but the earlier figures of 7 and 8 millions for the newborn child have not been confirmed by later work. Destruction of a large proportion of the extra cells occurs within the first ten days, but for a few days the red cell count shows a progressive fall, for some weeks after this the count remains considerably above that of the adult.

The destruction of large numbers of erythrocytes shortly after birth and, as a consequence, the liberation of excessive amounts of hemoglobin into the plasma has been held to be the cause of the physiological jaundice of the newborn, in whom the bilirubin content of the plasma (ch 40) is from 3 to 5.5 mg per cent. The infant's red cells are said by some to be more fragile than those of the adult, yet it is questionable, in view of the revised figures for red-cell counts in the newborn, that excessive hemolysis plays a very important part in the production of the jaundice. Again, others deny that the red-cells at birth are unduly fragile. Immaturity of the excretory function of the liver probably plays the dominant role.

The total number of red cells in the human body estimated upon the basis of 5 million per cubic millimeter is about thirty-three million million (33,000,000,000,000) which gives a total red-cell area exposed to the plasma of between

³ Though very widely accepted, this sex difference has probably been overemphasized. King and his associates found a maximal difference of only 1 per cent in favor of men.

3500 and 4000 square meters. Eamons has made the interesting observation that in the species which he investigated (cat, rabbit, dog and man) the surface area of the individual corpuscles and the corpuscular count varied in opposite directions, so that the cell area per cubic millimeter of blood was nearly the same in all—7.24, 7.55, 7.52 and 7.35 square centimeters for the four species in the order given above. The quantity of hemoglobin per square centimeter of red cell area is also about equal in all mammalian species.

PHYSIOLOGICAL VARIATIONS IN THE NUMBER OF RED CELLS

Increase in the number of red cells occurs under the following conditions:

(a) **HIGH ALTITUDES** It was first shown by Viault in 1889 that the inhabitants of mountainous regions, especially where the elevation above the sea is 10,000 feet or more, have constantly a much higher red cell count than persons living at sea level. The natives of some regions in the Peruvian Andes, where the altitude is 14,000 feet or more, have a red cell count 30 per cent above the normal (over 7 million per cubic millimeter). Not only the natives, but travellers sojourning even for a short time at these altitudes undergo an almost immediate increase in the number of their red cells. The corpuscular increase is directly proportional to the altitude, as may be seen from table 3.

It is perhaps necessary to point out here that an increase or a decrease in the red cell count does not of itself inform one of an increase or decrease in the red cells of the body as a whole. The red cell count gives only an estimate of the *number of cells per unit quantity of blood* (p 17). A reduction in the amount of plasma or of the water of the blood, for

TABLE 3

ALTITUDE IN THOUSANDS OF FEET	CORPUSCLES IN MILLIONS PER CUBIC MILLIMETER
0.7	4.5
4.4	5.2
12.0	6.8
15.6	7.8
18.2	8.3

From Barcroft after Hingston. In this table the altitude (in thousands of feet) multiplied by 0.225 gives a figure which approximates the increase in red cell count (in millions per cubic millimeter) above that at sea level. In this instance the count at sea level was 4.25 millions.

instance, would cause the *proportion* of red cells in the specimen to be increased. But neither the immediate nor the permanent increase in number of red cells at high altitudes is due simply to a reduction in fluid and a greater concentration of the blood, for estimations of the total volume of blood in the body (see ch 3) prove that there is an *absolute* increase in the number of circulating cells. How is the *immediate* increase brought about? It is obviously impossible to account for such a rapid rise in the red-cell count by a greater production of cells by the bone marrow, and it is at once suggested that large masses of cells are packed away from the general circulation—in a storehouse of some kind—but are quickly mobilized upon demand. When the functions of the spleen are considered (p 69) it will be seen that this organ serves as a reservoir for red cells and is responsible for the sudden increase in their number which occurs early in the process of acclimatization to high altitudes.

The *permanent* and great elevation of the red cell count, which is a characteristic feature of the blood of natives and other persons after acclimatization to the rarefied atmosphere, cannot be explained in the same manner, for the number of cells which the spleen can put into circulation is limited. Under these circumstances there is, actually, an increased manufacture of erythrocytes by the blood-forming organ—the bone marrow. The cells which are formed by the over-stimulated marrow are discharged into the general circulation at a somewhat immature stage of their development. They are spoken of as *reticulated cells* or

reticulocytes since their protoplasm shows a delicate filigree or reticulum which stains with basic dyes (fig 2.5 and p 15). The ultimate cause of the corpuscular increase (a physiological polycythemia) is undoubtedly the lowered oxygen tension of the atmosphere, and consequently of the blood, since animals placed in an hermetically sealed cabinet and subjected to lowered oxygen tensions exhibit similar blood changes.

(b) **MUSCULAR EXERCISE** and certain **EMOTIONAL STATES** cause a temporary increase in the number of red cells as a result of an outpouring of concentrated blood from the spleen.⁴ This may be looked upon as an emergency measure and, like that which occurs at high altitudes, is the response of the body to the tissues' call for oxygen.

(c) **HEIGHTENED ENVIRONMENTAL** temperature also causes a liberation of red cells from the splenic reservoir.

(d) Other conditions which tend to lower the oxygen tension of the arterial blood cause a rise in the number of circulating red cells. As in the response at high altitudes two factors are concerned, e.g., a discharge of blood from the spleen and other reservoirs, and a greater production of cells by the bone marrow.

Reduction in the number of red cells occurs at high barometric pressures, e.g., when the oxygen tension of the blood is higher than the normal. Animals, for example, living in deep mines have a lower red cell count than those at sea level.

ALTERATIONS IN THE NUMBER OF RED CELLS IN PATHOLOGICAL STATES

Increase in the number of red cells—polycythemia

Increase in the total number of red cells occurs as a compensatory measure in several pathological conditions and then represents apparently the response of the bone marrow to low oxygen tensions in the arterial blood. A red cell concentration of 7 million or more per cubic millimeter of blood is not unusual in the following conditions:

(a) **EMPHYSEMA** (p 430) and other chronic diseases which interfere with the oxygenation of the blood in the lungs (anoxia), e.g., tracheal stenosis, pneumothorax, tumor of the lung, pulmonary tuberculosis and pulmonary arterio-venous aneurysm.

(b) **CONGENITAL HEART DISEASE** (p 432).

(c) **AYERZA'S DISEASE**, a condition associated with

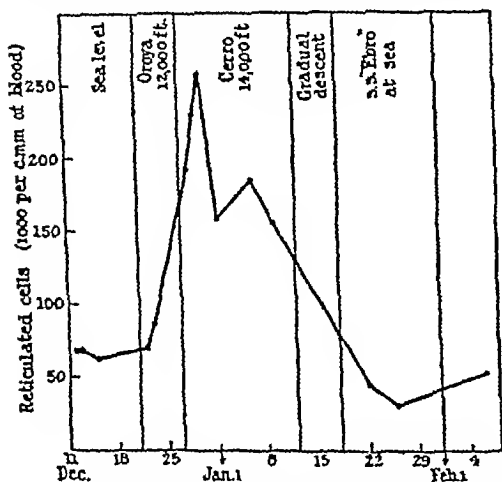


FIG. 2.5 Showing reticulocyte response to altitude (after Barcroft)

⁴ Though this can be demonstrated in experimental animals, it does not appear to be an important feature in man.

dilatation and marked hypertrophy of the right heart, sclerosis of the pulmonary arteries and their branches with consequent obstruction of the blood-flow through the lungs. There is extreme cyanosis, emphysema, dyspnea, and attacks of asthma are common accompaniments. The bone marrow is hyperplastic.

(d) CHRONIC CARBON MONOXIDE POISONING

(e) CHEMICALS, e g, chronic poisoning with arsenic, phosphorus and manganese, gum shellac, and certain aniline dyes

(f) REPEATED SMALL HEMORRHAGES, the polycythemia then represents an over-response of the bone marrow to the successive blood losses

POLYCYTHEMIA VERA (synonyms erythremia, splenomegalic polycythemia, Vasquez-Osler disease) Unlike the preceding types the polycythemia is not secondary to any known pathological condition. The disease appears to be primarily centered in the red bone marrow which is greatly increased in amount and hyperplastic, extending into the shafts of the bones to displace the fatty marrow, and packed with normoblasts and mature erythrocytes. Megaloblasts are absent or very scarce. The number of red cells in the circulating blood may be as high as 14 million per cubic millimeter and a count as high as 20 million has been reported. The total blood volume is greatly increased. The viscosity of the blood is of course greatly elevated (p 149). The concentration of hemoglobin in the individual cells and the chemical and physiological properties of the pigment are normal. The size, shape and general features of the red cells as a rule show nothing unusual and the number of reticulocytes is not greatly increased.

The chief features shown by the disease apart from those of the blood itself are cyanosis (p 435), dyspnea on exertion, enlargement of the spleen, hemorrhages and a familial tendency. Death may occur from thrombosis of the portal vein or a cerebral vessel. The circulation rate (ch 26) is slowed and the diffusion rates of oxygen and CO_2 (ch 31) through the pulmonary epithelium are reduced, though during rest the oxygen saturation of the arterial blood is usually normal. The renal blood flow and the filtration fraction (chap 35) are increased whereas the plasma flow is decreased. The oxygen saturation of the blood is reduced, however, during exercise. Owing to the great increase in hemoglobin concentration the actual *quantity* of oxygen in the blood is greater than normal. If the polycythemia were a compensatory reaction brought about by the lowered rate of diffusion of oxygen through the pulmonary epithelium, one would expect that breathing air with a high pressure of oxygen would be of benefit, but this is not the case. Exposure in a chamber to a high oxygen tension for several

days does not effect a reduction in the number of red cells. The reduced circulation rate may result in some way from the fact that the blood contains such a large load of oxygen that the tissues can obtain their quota from a smaller quantity of blood than normally. On the other hand, the reduced circulation rate may be due primarily to vasoconstriction in some part of the circulation. Some believe that narrowing of the caliber of the vessels of the bone marrow and the resulting low oxygen tension produced thereby provide the stimulus for the overproduction of blood cells. Studies of blood lactic acid concentration in polycythemia vera following muscular exercise lend support to the idea that a sluggish blood flow through the tissues is a causative factor. In normal persons and in the secondary types of polycythemia mentioned above, a rise in blood lactate occurs, whereas in polycythemia vera exercise causes a fall. This anomaly could be explained upon the basis of a high resting blood lactate as a result of a slow blood flow and the accumulation of lactic acid in the ischemic tissues. Exercise would then, by causing vasodilatation and a freer oxygen supply, in the contracting muscles, tend to reduce the concentration of the metabolite in the blood. That this is the probable explanation is indicated by the observation that vasodilatation induced by heat also causes a fall in blood lactate in this disease. The phenomenon appears to be of fundamental significance, and not simply the result of the high erythrocyte concentration, because it is not abolished when the red cell count is brought down to normal by treatment with *phenylhydrazine hydrochloride*, a drug which has been employed in controlling the disease.

Polycythemia can be produced in dogs by the daily administration of cobaltous chloride (8 mg daily for 2 or 3 weeks). The high red cell count thus induced is reduced to normal by feeding with whole beef or hog liver or by the daily injection of ascorbic acid. The effect of liver suggests the presence of an hepatic hormone possessing a depressant action upon bone marrow activity. Radioactive phosphorus (P^{32}) has been used lately with success in the treatment of the disease.

The observations of Schafer have thrown a new light on the pathogenesis of polycythemia vera which may show a way to the rational treatment of the disease. In experiments upon dogs he found that excision of the carotid sinus and the aortic nerves caused the development of polycythemia and hypertension in 40 per cent of the animals. Removal of the paravertebral chain of sympathetic ganglia was followed by a gradual return of the erythrocyte count to normal. A similar result followed paravertebral sympathectomy in a patient suffering from polycythemia. The polycythemic response to removal of the influence of the carotid sinus and aortic nerves is possibly due to constriction of the vessels of the bone marrow, the anoxia caused, thereby acting in the usual way to stimulate its hemopoietic function.

Reduction in the number of red cells below the normal is known as anemia. The causes and varieties of anemia are manifold and will be considered in chapter 9.

An apparent decrease or increase in the number of red cells occurs under certain conditions which upset the water balance of the body. It has already

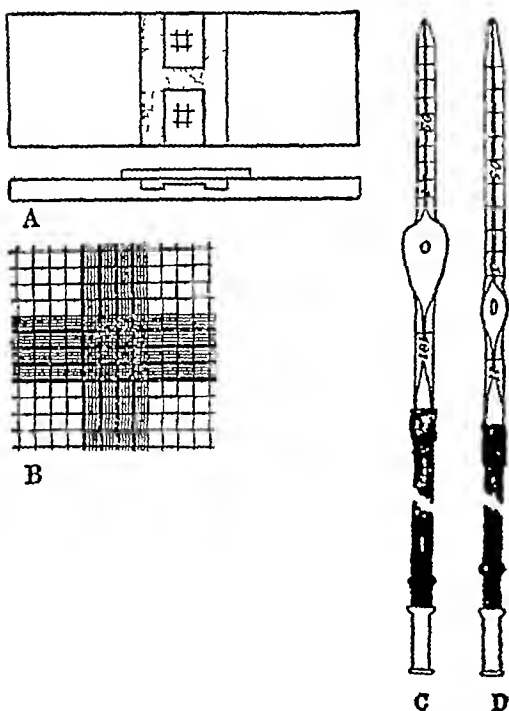


FIG 26 Hemocytometer. A, Counting slide shown on the flat and in cross section, it consists of two platforms surrounded on three sides by a trench, scale is engraved in the center of each platform. B, scale enlarged, C, pipette for diluting the blood 1 in 200, in counting of red cells, D, pipette for diluting the blood 1 in 20, in counting of white cells.

A drop of blood after dilution with a solution of 0.9 per cent sodium chloride is placed upon each of the glass platforms and a specially made cover glass gently applied. The space, filled with the diluted blood, between the surface of the platforms and the cover glass is $\frac{1}{16}$ mm. The scale is marked off in squares $\frac{1}{16}$ sq mm in area. Larger squares, each containing sixteen of the smaller squares, are marked off by heavier lines. The red cells in four large squares (on each platform) are counted under the low power of the microscope. The total number of cells is divided by one hundred and twenty eight ($16 \times 4 \times 2$) which gives the average number in one small square. Now, the depth of the fluid is $\frac{1}{16}$ mm, so the volume upon each small square is $\frac{1}{16} \times \frac{1}{16} = \frac{1}{256}$ cu mm. If the average number of red cells in each square is 6, then since the dilution was 1 in 200, the number per cu mm of blood is $6 \times 4000 \times 200 = 4,800,000$. The method of counting the white cells is similar in principle, but the blood is diluted 1 in 20 with a fluid which destroys the red cells and stains the white cells. The error in counting the red cells, even by the most experienced, is around ± 7 per cent.

been pointed out that the red cell count gives merely an estimate of the proportion of cells to plasma. There may, for instance, be undue retention of water in the body, and the plasma or its watery constituents may then be increased (hy-dremia or hemodilution). The blood is diluted, and the number of red cells per unit of blood is reduced, yet there is no absolute decrease in the number of circulating cells. Conversely, a loss of plasma or of merely the water of the blood (anhydremia or hemoconcentration) will increase the red cell count, i.e., the blood becomes more concentrated though the total number of red cells in the body is not altered. Therefore in conditions associated with extreme dehydration (p. 24) of the body the number of red cells per unit of blood is increased.

A practical point to be remembered when counting the red cells or in determining the packed cell volume is that their number may be relatively increased or decreased by purely local alterations in blood concentration. Pronounced dilatation of the capillaries of the region of skin from which the sample has been taken will cause a local slowing of the blood stream and congestion of the part. The loss of fluid from the vessel into the tissues and the clumping of corpuscles which may result from the greater capillary pressure will give a false estimate of the number of red cells. On the other hand, pressure made upon the part by the examiner in order to hasten the flow of blood from a skin puncture may express fluid from the tissues which will dilute the red cells in the specimen. Moreover, even under ordinary physiological conditions, the concentration of red cells may vary considerably in different parts of the circulation. The proportion of red cells in the capillaries is greater as a rule than in the heart and larger vessels, by around 12 per cent, but in certain states associated with slowing of the peripheral circulation, concentration of red cells in the capillaries may increase the disparity. Trapping of plasma in the peripheral vessels may occur in other conditions and will tend to reduce the concentration of cells in the capillary areas below that in the general circulation. These are also important points to bear in mind in estimating the blood volume by the dye method (p. 18).

Estimation of red cells. The concentrations of red (or white) cells may be obtained by counting them directly after suitable dilution beneath the microscope, as in the method of Thoma-Zeiss. The instrument used for this purpose is called a hemocytometer. The reader is referred to texts on laboratory methods for details (see also fig. 26). Another method is by means of the hemato-crit (fig. 27). In this the plasma and corpuscles are separated by centrifugal force. The blood, rendered non-coagulable, is drawn into a graduated capillary tube, placed in a centrifuge and revolved at a speed of 3000 revolutions per minute for from 30 to 60 minutes. At the

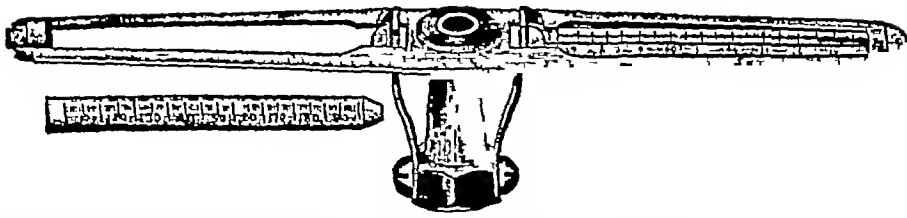


FIG 27 Hematocrit (Daland) The graduated glass tubes are filled with blood and placed in the carrier which is then rotated on a centrifuge (see text)

end of this time the original blood will be found to have separated into a clear colorless column of plasma and a red column—the corpuscles. The lengths of the two columns are read off by means of the graduations on the tube. The normal proportions of plasma and corpuscles in human blood are about 54 (53–56) and 46 (44–47) respectively. That is, the volume of cells (*packed cell volume*) is 46 per cent of the total volume of the specimen of blood. The average volume of the individual erythrocytes is obtained by dividing the packed cell volume by the figure for the red-cell count.

VARIATIONS IN SIZE, SHAPE AND STRUCTURE OF THE RED CELLS

Under physiological conditions little change in shape of the red cell occurs, though a few fragmented cells may be found in normal blood. These are, as will be seen later (p 73), simply remnants of senile cells which have undergone a natural disintegration in the blood stream. A slight change in volume, about 7.5 per cent occurs, due to osmotic changes incident to the respiratory cycle (p 397). The red cells at birth and in early infancy are larger than in adult life. This as well as the higher cell count is responsible for the high packed cell volume (54 per cent) of the infant's blood.

In *disease*, the red cell is subject to many changes in size, shape and structure. The examination of the blood and the identification of the various forms of abnormal cells is an important means employed in diagnosis of the different anemias (p 80). Only the more outstanding abnormalities can be touched upon here. The least pronounced departure from the normal blood picture is an increase in the number of *reticulocytes* (see frontispiece). These young red cells resemble the ordinary cells in every way except that after supravital staining (alcoholic solution of brilliant cresyl blue added to blood in the fresh state) a fine reticulum of basophilic material can be seen in the cytoplasm. The reticular material is of cytoplasmic origin, probably gamma-globulin, and does not represent nuclear remains. A reticulum of similar nature may sometimes be seen in cells in which the nucleus

is still intact. In normal human blood the reticulocytes are from 0 to 2 (av 0.8) per cent of the total red cell count.⁶ They are increased after hemorrhage, at high altitudes (p 11), by exercise, in acholuric jaundice (p 80) and in pernicious anemia, especially following specific treatment (p 84) of the latter condition and during the blood crises. A rise in the reticulocyte count indicates an increased activity of the blood forming tissue—the bone marrow—which as a result of a specific stimulus turns out a larger number of young cells. The maturation of the red cells, that is, the change from reticulocytes to erythrocytes, has been estimated to take from 10 to 24 hours. There is evidence that the thyroid gland liberates a principle which hastens the maturation of reticulocytes.

The next stage in the life of the red cell is represented by the *normoblast*, which appears in the blood in several different types of anemia. This cell, as its name implies, is normal in size and shape, possesses the usual amount of hemoglobin, but contains a nucleus. The bone marrow normally holds large numbers of these immature cells, but in health they do not reach the general circulation. In pernicious anemia and other severe types of anemia large, pale, nucleated cells are seen resembling white cells (lymphocytes, p 92). They represent a very primitive stage in the development of the red cells, they are termed *megaloblasts*. The megaloblast contains a small, sometimes a negligible amount of hemoglobin. Its protoplasm also contains a diffuse or punctate arrangement of basophilic material. Hemoglobin is acidophilic, so these cells may stain with acid as well as with basic dyes. This phenomenon of dual staining, which may also be shown by other abnormal cells, is known as *polychromasia*. Other cells of various sizes and shapes may be seen in the anemias. Shrinkage of the contents of the red cell with wrinkling of its limiting membrane, as may

⁶ These cells are more numerous in rabbit's blood, being from 3 to 3.5 per cent of the total red cell population. The blood of certain other species, also shows a relatively high count.

result from immersing the cells in a hypertonic solution, is termed *crenation*. In pernicious anemia, in particular, the presence of cells of unequal sizes (*anisocytosis*) and of deformed outline (*poikilocytosis*) is common. The poikilocytes may assume the most bizarre forms, mulberry, flask or hammer shapes may appear. Erythrocytes with a globular form (*spherocytes*), or of an elliptoid or crescentic shape (*elliptocytes* and *sickle cells*) are found in certain types of anemia. Such cells are more fragile than normal and, owing to their being less prone to form rouleaux, the blood undergoes sedimentation (p. 66) more slowly. *Macrocytes* and

microcytes are terms denoting cells of usual structure and without nuclei, but larger and smaller respectively than the normal erythrocyte. In certain conditions, e.g., lead poisoning, fine dots of basophilic material, probably a porphyrin (ch. 6) appear throughout the cell, giving it a stippled appearance. This abnormality is known as *punctate basophilia* (see frontispiece). In certain types of anemia rings or twisted strands of basophilic material may be seen near the periphery of the cell. These are derived from the nucleus and are known as *Cabot's rings*. At other times small nuclear fragments—*Howell-Jolly bodies*—are present in the cytoplasm.

CHAPTER 3

THE BLOOD VOLUME BODY WATER WATER BALANCE

It has already been pointed out that a red cell count is merely an index of the relative proportions of cells to plasma and does not permit one to say whether the body's total supply of erythrocytes is increased or diminished. The same is true, obviously, for a hematocrit determination. These two methods indicate the *concentration* of the blood in erythrocytes. As a general rule, however, since the volume of the blood tends, through changes in its fluid content, to regain its normal level after this has been reduced or increased, the concentration of red cells *does* parallel the body's total supply. Yet it will be clear from the following examples that the principle upon which these methods are based does not imply any such relationship. If an erythrocyte count be made immediately after a sudden hemorrhage, it will be found to be practically normal, since by this time little or no fluid has passed from the tissue spaces into the blood vessels to replace the lost blood, that is, to dilute what remains in the vessels. On the other hand, if a quantity of fluid be injected into the veins of an animal the blood count will be reduced, yet no reduction in the total number of cells has occurred. Had there been no previous knowledge of the conditions present in either of these instances, the red cell count would have given quite erroneous information regarding the body's total supply of erythrocytes. This may all seem very obvious to the reader, yet before clinical methods had been devised for the measurement of the blood volume, many wrong conclusions had been drawn concerning certain types of anemia. Blood volume estimations are of value in many clinical conditions associated with a loss or gain of fluid by the body, or for the purpose of checking the results of corpuscular and hemoglobin estimations. It is also sometimes of interest to know, in experimental investigations and in metabolic studies in man, the total amounts of certain blood constituents e g, protein, calcium, sodium, etc., as well as their concentrations.

METHODS FOR THE ESTIMATION OF BLOOD VOLUME

DIRECT METHOD

The first attempts to measure the total quantity of blood in the body were made upon animals by Welcker

(1854). His method consisted in taking a small measured quantity of the animal's blood and diluting it to 1 in 100 with normal saline.

The animal was then bled, and after the blood had ceased to flow, its vessels were washed out and the muscles minced and extracted with water. Water was then added to the collected fluid—blood and washings—until its color matched precisely the tint of the original diluted blood specimen. The total collected fluid divided by 100 gave the blood volume.

This method was also employed upon decapitated criminals (Bischoff) in order to obtain a value for the blood volume of man. By this method the total blood of the body was found to be about $\frac{1}{15}$ of the body weight—that is, from 5 to 6 liters (9–11 pints) in a man of average weight.

Obviously Welcker's method is applicable only to physiological problems of the laboratory. Furthermore, it is not free from serious errors: (1) The coloring matter of the muscles interferes with the calculations; (2) The turbidity of the final solutions prevents exact color matching; and (3) the loss of blood pigment as a result of intravascular as well as extravascular clotting is considerable and leads to further error.

INDIRECT METHODS

Carbon monoxide method

For the determination of the blood volume during life one or other of two methods may be employed. The first of these is the carbon monoxide method originally devised by Grehan and Quinquaud for animals, and modified by Haldane and Smith for man. The principle of this method is based upon the following facts: (a) carbon monoxide when it combines with the blood pigment (hemoglobin) gives a bright cherry red color to the blood, so that the amount of the compound formed may be estimated colorimetrically; (b) carbon monoxide displaces oxygen from hemoglobin volume for volume.

The method is carried out as follows—the *oxygen capacity* of the subject's blood, that is, the maximum amount of oxygen which 100 cc. of his blood will absorb—is first determined. The oxygen capacity of normal blood is around 20 cc. Since the volumes of CO and O₂ that a given quantity of hemoglobin will take up are equal this determination gives the CO capacity as well. The subject is then instructed to breathe a known amount of CO (say 110 cc.) from a bag. Subsequent color comparisons of the blood with a standard color solution show that the absorption of this quantity of gas has satisfied the blood, say, to only 12 per cent of

its capacity. It is a simple matter of calculation to determine from this datum how much gas would be required to saturate the blood to its full capacity, and so arrive at the total blood volume. Thus—

After the inhalation of 110 cc. of gas the blood is found to be 12 per cent saturated, therefore $110 \times (100/12) = 916$ cc. of gas would be required to saturate it to full capacity. Inasmuch as each 20 cc. of gas would represent 100 cc. of blood if this were fully saturated $(916/20) \times 100 = 4580$ cc. total blood volume.

The carbon monoxide in the blood may be estimated more accurately by gas analysis after the method of Van Slyke rather than colorimetrically. The technical difficulties of the carbon monoxide method of determining blood volume are considerable especially in persons who are very ill.

The dye method

This method was originally devised by Keith, Rowntree and Geraghty and has largely supplanted the carbon monoxide method. The degree of dilution in the plasma of a known amount of dye injected into the circulation is employed as the basis of calculation. The color of the stained plasma is compared in a colorimeter with that of a standard dye solution of known concentration. There are several qualifications required of a dye before it can be considered suitable for blood volume measurements. In the first place, of course, it must be innocuous. It must not diffuse too rapidly from the blood stream, it must color only the plasma, and not be adsorbed by the cells of the blood nor by the walls of the blood vessels. Finally, it must not change color itself after entering the blood, nor cause the liberation of pigment from the red cells (hemolysis). If it lacked these qualifications, the colorimetric readings obviously would be undependable.

The dye most commonly employed until recent years was brilliant vital red. The method has been improved by Gregersen and his associates by the use of the blue dye T-1824 (Evans blue). Evans blue has the advantage that the error due to any discoloration of the plasma by hemoglobin (hemolysis) is minimized. This dye is also eliminated very slowly (4.8 per cent per hour) from the circulation. The color determinations are made most satisfactorily with a photoelectric colorimeter or a spectrophotometer. In the method as modified by Crooke and Morris, the proteins of the plasma samples are precipitated and removed by filtration and the colorimetric measurements made upon the filtrate. The following is the technic employed by these authors.

Six ml. of blood are drawn from a median antecubital vein for making up a standard solution. Five ml. of a 0.70 per cent sterilized Evans blue solution are injected through the same needle. At intervals of 20, 40 and 60 min., 3 ml. samples are drawn from the opposite antecubital vein. All samples are transferred to 15 ml. centrifuge tubes, which have been coated with paraffin wax, and which contain 30 mg. of a 2:3 mixture of potassium and ammonium oxalates. After mixing, the tubes are centrifuged for 10 minutes at 2500 revolutions per minute. The plasma is immediately removed and kept in stoppered tubes for analysis. The proteins of the plasma samples are precipitated by adding slowly 7 ml. of the precipitating reagent (1 vol. concentrated HCl and 6 vol. alcoholic phosphotungstic acid) to 1 ml. of plasma. The samples are centrifuged and the supernatant fluid removed for colorimetric determinations. A standard solution is made up by adding 0.10 ml. of a 1 in 50 dilution of the injection solution to 1 ml. of undyed plasma. Of this mixture, 1 ml. serves as a standard. This quantity of the standard solution contains 1/2500 of the total amount of dye injected, such a dilution corresponds, therefore, to a "plasma volume" of 2500 ml. A color measurement is then made on each sample.¹

The calculations are made as follows. When P is the plasma volume in milliliters, V_1 the mean value of the photometric or colorimetric readings of the three plasma samples and V_2 the value for the standard, then

$$P = \frac{V_2 \times 2500}{V_1}$$

In order to obtain the value for the total blood volume the proportion of red cells to plasma must be determined by means of the hematocrit. If the packed cell volume is say 45 per cent, the plasma volume is, therefore, 55 per cent and the total plasma volume, as determined above, is, say $(3000/55) \times 100 = 5454$ cc. total blood volume.

The use of radioactive iron to estimate the circulating red-cell volume

In calculating the blood volume from the plasma volume and the hematocrit reading it is necessarily assumed that the proportion of red cells to plasma is the same throughout all parts of the vascular system, but experimental evidence is strongly against such an assumption. In capillaries the proportion of red cells to plasma may be considerably less or more than in the

¹ There has been some question whether a single injection of T 1824 gives reliable results. It has been stated that the dye is taken up by the reticuloendothelial cells and that it is necessary to "block" this system by a preliminary injection, otherwise the estimated plasma volume would be too high. On the other hand, a second injection would not be taken out of circulation and would all be mixed with the plasma. This blocking effect was discovered in the cat ("cat effect") but it cannot be demonstrated in man.

larger vessels. To eliminate the error of unequal distribution of red cells throughout the circulation, a method has been devised by Hahn and his associates to estimate the total *circulating red cell volume* by means of an isotope of iron. This was later modified by Gibson and his colleagues.

Radioactive iron (Fe^{55} or Fe^{59}) is incorporated into ferric ammonium citrate, which is injected into a donor belonging to group O (p 42). Newly formed cells take up the iron and appear in the circulation in 24 hours, reaching a maximum in 21 days.

From 70 to 100 milliliters of the donor's blood (30 to 40 milliliters of red cells), having a radioactivity of from 2500 to 3000 counts per minute per milliliter (determined by Geiger counter), are injected intravenously or intra-arterially into the subject. The quantity of donor's blood injected will depend upon the volume of red cells which the recipient is expected to have, but injections between 70 and 100 milliliters cover a range of red cell volumes of from 1500 to 2500 milliliters.

Between 10 to 20 minutes are allowed to pass after the injection, in order for thorough mixing of donor's and recipient's cells to occur. Samples of 15 ml of recipient's blood are then taken and are followed at 10 and 20 minute intervals by two more samplings. Two ml of donor's blood are diluted to 100 ml, 10 aliquots are then prepared. The recipient's blood samples are centrifuged at 3000 r.p.m. for 30 minutes. Donor and recipient samples are *wet-ashed* and the iron deposited electrolytically on copper. The radioactive iron in the two samples is determined by means of a Geiger counter, the calculation of the circulating red cell volume is made from the following equation:

$$V_{rr} = \frac{CD \times VaD}{VaR},$$

where V_{rr} is the red cell volume, CD the number of milliliters of donor cells injected, VaD the radioactivity of the donor's cells and VaR the radioactivity of the recipient's cells.

In order to determine the total blood volume, the plasma volume is measured by the dye method and the result added to the value for the circulating red-cell volume. The blood volume estimated in this way agrees well with the CO method but is lower than that arrived at by the dye method.

This method is based on two assumptions—(a) that none of the radioactive iron escapes again after being taken up by the erythrocytes, and (b) that all the labelled cells became mixed with the recipient's cells equally throughout the circulation. Only the principle of the method can be given here, for details the reader is referred to the original papers.

All determinations of blood volume, whatever method is used, are made with the patient recumbent and under basal conditions, some 12 to 14 hours after a meal. The

loss of dye into the tissues at the site of injection must be avoided since this of course will vitiate the results which, since the plasma would be less deeply stained, would be too high. Care must also be exercised to prevent evaporation of fluid from the samples after they have been drawn. Otherwise the concentration of dye in the plasma would be raised and the readings (which would indicate a lower degree of dilution) would be too low. Determinations by the dye method from time to time on the same individual under comparable conditions give consistent results. The absolute blood volume of course will never be known exactly, since it cannot be measured directly in the living subject. But even in the dead subject there is no reason to believe that the direct (Welcher's) method is any more accurate than the dye method in the living.*

NORMAL VALUES FOR PLASMA AND WHOLE BLOOD VOLUMES AS OBTAINED BY THE DYE METHOD

The whole blood is about $\frac{1}{12}$ and the plasma $\frac{1}{8}$ of the total body weight, i.e., 8 and 4 per cent respectively. Expressed as volumes, the whole blood is about 78 cc, the plasma volume about 41 cc, and the red cell volume about 37 cc per kilogram. The blood volume of a man of average weight (70 kg) is therefore around 5500 cc. Rowntree has demonstrated that the blood volume is a function of the body's surface area. It amounts to about 3 liters per square meter (p 618). The blood volume per square meter is higher by about 75 per cent in men than in women, the plasma volumes in the two sexes are, however, about the same. That is, the larger volume in males is attributed to a greater number of red cells (see p 11). In infants and young children the blood volume is somewhat less per kilogram of body weight and per square meter of body surface than in adults, but gradually increases throughout childhood to reach the adult figure at around 16 years of age.

BODY WATER (see also ch 4)

The blood volume and its variations cannot be considered entirely apart from the fluid content of the body as a whole. Blood volume regulation is largely a question of balance between the fluid within the vessels and in the tissues. When conditions arise which tend to lower or raise the volume of blood, counter forces (p 31) come into play.

*The blood volume as estimated by the foregoing procedures gives the volume of circulating blood and does not include the blood of the spleen (p 69) which plays the rôle of a reservoir.

which restore the normal level. When circulating fluid is lost, the vessels replenish themselves from the extravascular spaces. On the other hand, any tendency for the blood volume to rise is met by a discharge of the excess fluid into the tissues and later from the body in the urine. So, a balance is struck, and in health the blood volume is maintained remarkably constant. For example, after the intravenous injection of a large quantity of saline, the volume of circulating fluid (though raised temporarily) is brought back to normal within 30 minutes or less. On the other hand the loss of blood fluid, as by hemorrhage, immediately calls into action processes which may, in a very short time, replenish the blood volume. When an animal is bled to death, at a not too rapid rate, the blood which is withdrawn is found after a few minutes of bleeding to have become diluted—clearly demonstrating the promptness with which fluid (water and salts) has been absorbed into the vessels.

The total volume of the water of the body varies among species but in man has a mean value of 62 (52 per cent for women)³ per cent of the body weight (it is about 65 per cent in the dog and about 73 per cent for guinea pigs and rabbits). The total volume of the body water can be determined by injecting into the blood stream a known amount of heavy water, D₂O, or antipyrine which becomes uniformly distributed throughout the body fluids (intracellular and extracellular), and then determining its concentration in a sample of serum.

The water content of various tissues, in average percentages, is given in the following table.

	per cent
Muscle (striated)	75
Skin	70
Connective tissues	60
Adipose tissue	20
Bone (marrow-free)	25-30
Blood	
Plasma	90
Cells	65
Kidney	80
Liver	70
Nervous tissue	
Gray matter	85
White matter	70

³ The total body water is related to the mass of the lean tissues of the body rather than to the body weight. In an obese person it is a much lower percentage of the body weight than in one who is lean. The values, therefore, vary widely both between persons and different species of animal in accordance with their fatness or leanness.

The *extracellular fluid*, which comprises the blood plasma, the tissue or interstitial fluid, lymph and the fluid in the serous cavities, amounts to about 20 per cent of the body weight, the plasma water constitutes only about 4 per cent of the body weight. The fluid within the cells, the *intracellular fluid* amounts to 45 per cent of the weight of the body, or over 2 times the extracellular fluid. The skeletal muscles contain about half, the skin about $\frac{1}{3}$ and whole blood only about $\frac{1}{10}$ of the total body water.

In general, the intracellular fluid has a high concentration of potassium and a low concentration of sodium, whereas the extracellular fluids contain relatively large amounts of sodium and small amounts of potassium.⁴ In most cells chloride is also in low concentration or is absent, the red blood cells and the cells of the gastric glands are notable exceptions. Substances, such as sucrose, mannitol, inulin, thiocyanate and to a large extent chloride, when introduced into the body, become uniformly distributed throughout the extracellular fluids, but do not enter the cells in important amounts and, since they are not metabolized and are excreted not too rapidly, can be employed to estimate the volume of the extracellular fluid in the living animal. When, for example, a known amount of thiocyanate, is injected into the blood stream and sufficient time allowed for equal distribution to occur, the volume of the extracellular water can be calculated from the concentration of the solute in a sample of serum. Radioactive sodium or chloride may also be employed for such determinations, but since these solutes are not excluded from all types of cells they are not as dependable as thiocyanate or sucrose, especially if absolute values are sought. The extracellular fluid is visualized as occupying a space of the determined volume, and according to the solute used in the estimation it is customary to speak of the *thiocyanate space*, *sucrose space*, etc.

The values for the total body water and the

⁴ In the case of muscle at any rate, this distribution of Na and K between intra- and extracellular fluids may be altered in pathological states involving the excessive loss of potassium from the body, e.g. severe and prolonged diarrhea, hyperactivity of the adrenal cortex or the administration of desoxycorticosterone (ch. 59) when muscle potassium is largely replaced by sodium. Increased concentration of K in the serum and other extracellular fluids occurs in adrenal cortical insufficiency, in oliguria or anuria, during tissue breakdown or anoxia, following major surgical operations, and in dehydration.

extracellular water being known, the intracellular water is got by subtraction (see table below)

Body water	
BODY WATER	PERCENTAGE OF BODY WEIGHT
A Extracellular	
plasma	4 0
interstitial	15 0
B Intracellular	45 0

The characteristic distribution of sodium and potassium between intracellular and extravascular fluid makes it possible to determine the proportions of excreted water derived from intracellular and extracellular sources, respectively (see table 4) The Na and K contents of the water excreted in certain pathological states is therefore employed as a means of estimating the proportions derived from extracellular and intracellular sources respectively Thus, during the subsidence of nephritic edema, or of the subcutaneous accumulations in myxedema following thyroid treatment, the excreted water contains a large excess of sodium over potassium During the fluid loss which results from a diet of glucose and water, the loss of extracellular fluid as well as the shrinkage of the tissues causes an increased excretion of both Na and K (Byrom) In the early stages of a fast the excreted water has a high sodium content, after prolonged fasting its relatively high potassium content indicates an intracellular origin (ch 51) A loss of one-fourth of the body water is usually fatal

TABLE 4

Concentrations of base in the water of blood plasma and in the water of muscle tissue
(After Gamble, Ross and Tisdall)

	PER 100 CC PLASMA	PER 100 CC PLASMA WATER	PER 100 CC MUSCLE TISSUE	PER 100 CC MUSCLE WATER
	mg	cc 0.1 N	mg	cc 0.1 N
Na ⁺	330	157.7	80	45.8
K ⁺	20	5.6	320	108.0
Ca ⁺⁺	10	5.5	8	5.3
Mg ⁺⁺	3	2.7	21	23.0
Total		171.5		182.1

Water of blood plasma is taken as 91 per cent by volume and water of muscle tissue as 76 per cent of weight.

Under physiological conditions the interstitial fluids show the greatest changes, the volumes of plasma water and intracellular fluid remaining relatively constant In dehydration the proportion of the total water from extracellular and intracellular compartments varies with the manner in which the negative water balance is produced When, for example, the dehydration is caused by sucrose diuresis, about 85 per cent of the excreted water is derived from extracellular and 15 per cent from intracellular sources Whereas, in dehydration resulting from water deprivation, the water loss is from 57 to 67 per cent intracellular (Painter and associates) As mentioned above, the source of the excreted water is indicated by its concentration in K or Na

Hyaluronic acid Though practically none of the extracellular water is chemically bound, that is to say, substances are dissolved in it readily, and it can be entirely separated from colloidal materials by ultrafiltration, and though free movement and rapid changes of distribution are permitted, its physical state does not appear to be that of a simple solution of electrolytes and protein bathing the cells, it is held in the tissue spaces by a gelling substance The latter has been identified chemically as a mucopoly saccharide and called *hyaluronic acid* This substance also enters into the formation of the cement substance binding cells together, and into the production of other gel-like materials, e.g., vitreous body, jelly of the umbilical cord, etc The physiological significance of this material is dealt with more fully in chapter 28

WATER BALANCE

In health, except when new tissue is being formed, the body's intake of water obviously must balance the output When the output exceeds the intake, the body's water content is reduced and the body is then said to be in negative water balance, dehydration results When, as during growth, convalescence from an acute illness or in pregnancy, new tissue is being formed, or for a time after a subject has been placed upon a reducing diet, the water balance is positive, the intake exceeding the output, i.e., water is retained

The antidiuretic principle of the hypophysis, and probably also an antidiuretic substance derived from the liver, constitute part of the mechanism regulating the volume of body water In dehydrated states, increased amounts of the pituitary hormone are excreted in the urine (ch 57) The hepatic material is vasodepressor as well as

antidiuretic in action. It is known briefly as VDM and appears in the blood of animals in the irreversible stage of hemorrhagic shock (pp 77 and 303) and in the circulation of patients suffering from cirrhosis of the liver, accompanied by increased plasma volume. This principle, which has been prepared in purified form from beef liver, exerts a powerful antidiuretic effect when injected intravenously into animals, inducing water retention and increasing the susceptibility to water intoxication (see Shorr and associates, and pp 27 and 77).

THE WATER INTAKE

Body water is replenished in two main ways, (a) by the ingestion of liquids, semisolid and "solid" food (cooked lean meat, for example, is from 65 to 70 per cent water), and (b) by the water formed in metabolism through the oxidation of the hydrogen of the food, or of the body tissues themselves. The following table from Rowntree compiled from the data of Magnus Levy gives the quantities of water produced by the metabolism, respectively, of the three main food stuffs and of alcohol.

100 grams of fat	yield 107	1 grams water
100 grams of starch	yield 55	1 grams water
100 grams of protein	yield 41	3 grams water
100 grams of alcohol	yield 117	4 grams water

Water is also formed in the tissues through the polymerization or synthesis of various compounds, i.e., through a metabolic process the reverse of hydrolysis. An ordinary mixed diet yields as a result of oxidative processes from 300 to 350 grams of water daily, or about 14 grams per 100 Calories. When no food or drink is taken, the body materials themselves are utilized for this purpose, the glycogen, protein and fat supplying important quantities of water. The camel's hump, for instance, which is largely composed of fat, is a reservoir for large amounts of water, and the clothes moth kept in a desiccator and fed upon perfectly dry food lays eggs which are 80 per cent water.

For the adult, the amount of water from all sources and under ordinary circumstances which must be ingested daily is around 2500 cc. or about 1 cc. per Calorie of food intake. This usually means that about 1000 cc. of water as such or in beverages must be drunk in order to maintain the water balance. Ordinarily, the volume of body water is held constant by varying the intake which

is regulated in turn by the sense of thirst. In experiments on dogs Robinson and Adolph found that when the animals had ready access to water they drank when they had lost water to the extent of 0.5 per cent of their body weight. The amount of water drunk was just sufficient to replace that which had been lost. It is hard to conceive what signaled the cessation of drinking, for it occurred while the water was still in the stomach, that is before it had been absorbed.

The water intake under conditions of average temperature, humidity and diet is summarized in the following table.

solid and semi solid food	1200
oxidation of food	300
drinks (water, milk, coffee, beer, etc.)	1000

THE WATER OUTPUT

Water is lost from the body in the feces, urine and saliva, and by the evaporation of water from the skin and lungs. The daily loss through these several channels is given in the following table for an averaged-sized man at light occupation in a temperate climate.

Skin (at average temperature and humidity)	500
Expired air (at average temperature and humidity)	350
Urine	1500
Feces	150
Total	2500

Under usual conditions of air temperature (23–25°C), humidity and diet, the heat lost from the lungs and the surface of the body by the evaporation of water amounts to about 24 per cent of the total heat production. The measurement of this *insensible water loss* under standard conditions may therefore be employed as a basis for the determination of the basal metabolism (p 617).

The loss in the saliva is negligible under ordinary circumstances but may be considerable in mouth breathing (as a result of evaporation) and in those addicted to the spitting habit.

The water lost through the skin and lungs varies greatly with the temperature and relative humidity of the atmosphere and with the extent of the muscular exercise indulged in. At ordinary temperatures slight secretion by the sweat glands is not perceived since the sweat evaporates as quickly

as it is formed. This *insensible perspiration*, as it is called, includes the loss of a greater amount of water by evaporation from the moist tissues beneath the skin, this loss is quite apart from the actual secretion of sweat (ch 53). The diffusion of water through the skin and evaporation from the surface under ordinary comfortable conditions of room temperature and humidity is around 1 mg per square cm of skin surface in a period of 10 minutes. The rate of diffusion is little different for living or dead skin and whether sweat glands are present or absent. The amount of the *insensible perspiration* has the average value given above but may be many times this value, when the air is hot or the body temperature raised, the rate of evaporation of water from the tissues beneath the skin is much more rapid, the secretion of sweat is also likely to be more active, but owing to the higher rate of evaporation a larger quantity of sweat is secreted before it becomes evident. Relative humidity and air movement also influence the rate of evaporation. So, in humid, still atmospheres sweat secretion is more evident though it may be no greater than in a drier atmosphere when evaporation is more rapid. Large quantities of sweat are secreted as a result of muscular exercise or when, as in the tropics, the temperature is high. In hot climates the daily secretion may amount to 3000 cc. daily and in very torrid atmospheres it may be as much as 10 liters. When heavy work is done in a hot environment, sweat may be secreted at the rate of 2 liters per hour. This necessitates the drinking of an equal quantity of fluid in order to maintain the normal water content of the body, since the intake must equal the output.

At ordinary temperatures the inspired air contains negligible quantities of water whereas the expired air is almost saturated with moisture. Any condition which increases the pulmonary ventilation therefore increases the water lost by this route.

The relation of the electrolyte concentration and tissue changes to the volume of body water

The isotonicity of the body fluid which depends mainly upon its concentration in sodium and chloride is maintained constant largely by the retention or elimination of water, the kidneys playing the primary role in this regulation. A loss of salt is accompanied by a loss of water and the ingestion of salt is followed by water retention. Thus it is possible to increase the volume of body water in normal persons to the point where edema occurs by the administration of large amounts of

sodium bicarbonate and to cause the discharge of nephritic or cardiac edema by the reduction or withdrawal of salt from the diet. In the latter instance, only sufficient sodium chloride is available for the production of a more limited amount of isotonic fluid.

Protein is laid down in the body with water (about 3 grams of water per gram of solid). During growth or convalescence from wasting diseases, retention of water therefore occurs, i.e., the intake of water, including that derived from solid food, exceeds the output. Fat is laid down with a minimal amount of water (only that in the protein of the connective tissue framework), the deposition of glycogen is accompanied by a small storage of water.⁵ Water retention therefore follows a sudden change from a high fat to a high protein diet, and to a less degree from a fat diet to one high in carbohydrate. A change from a diet high in carbohydrate or protein to one high in fat is followed by the loss of water.

Reduction of the caloric value of the diet below the energy requirement is accompanied for a time by the retention of water. During the first week or so on a reducing diet the subject's weight may for this reason show little or no change, the fat catabolized being replaced by water. In a prolonged fast after the fat stores have been depleted, protein is also partly replaced by water, the muscles of animals dying of starvation showing a marked reduction in the proportion of protein.

DEHYDRATION

When the output of water exceeds the intake, the body's water content obviously will be reduced. That is, the body is in negative water balance and the condition known as dehydration results.

Causes of Dehydration

Dehydration may result from

1 *Water depletion or primary dehydration* (a) *Simple deprivation of water* from whatever cause shipwreck, desert travel, dysphagia, extreme weakness, in mental patients who refuse to drink, etc. Under such circumstances, though there is an effort to conserve the stores of body water, through a reduction in the amount excreted by the kidney, in the sweat and by other routes, some water is always lost though none be drunk. Dehydration

⁵ Zuntz concluded from his experiments that 3 grams of water were laid down with each gram of carbohydrate, but his results have been questioned and are not now generally accepted.

occurs more quickly in fever or if the environmental temperature is high

(b) *Excessive water loss* may result from persistent vomiting (e.g., pyloric or intestinal obstruction) prolonged diarrhea, or the excretion of large quantities of urine or sweat, especially when accompanied by a restricted water and salt intake. In the acute diarrheas of infants, dehydration and loss of weight may occur very rapidly

In water depletion the osmotic concentration of the extracellular fluid rises, water is drawn from the cells, and both extracellular and intracellular compartments shrink. Extreme thirst is experienced

2 *Reduction in the total quantity of electrolytes, salt depletion or secondary dehydration* The electrolytic concentration of the body fluids, both extracellular and intracellular, is maintained constant through the elimination or retention of water. That is, a reduction or increase in the total electrolytes, which comprise chiefly the basic radicles Na (extracellular) and K (intracellular) and the acid radicles HCO_3 and Cl, is accompanied by a corresponding decrease or increase in the volume of body water. The sum of the basic elements and acid elements of course must balance. Loss of Cl can be made good by the retention of H_2CO_3 and a rise in plasma bicarbonate. Excreted base, however, can be replaced only by basic substances supplied in the food. The total concentration of electrolytes in the body fluids is therefore dependent upon the stores of total base. For example, in pyloric or high intestinal obstruction (p. 594) fluid is secreted in large quantities into the gastrointestinal tract. The fluid may be vomited or may collect and remain in the dilated part of the canal above the obstruction. (The latter occurrence is the rule in the rabbit which cannot vomit.) In either case the secretion of large quantities of gastric juice entails a loss of blood chloride. A similar chloride loss is induced in animals by means of a gastric fistula fashioned by sectioning through the pylorus, stitching the stomach opening to the abdominal wall and allowing the gastric juice to drain to the exterior. In the foregoing instances, the chloride depletion causes at first no ill effects, the normal concentrations in electrolytes of the blood and tissue fluids being maintained for a time by the retention of CO_2 , and, as a consequence of this, an increase in bicarbonate. The compensation for the Cl loss leads however to alkalosis which is then countered by an increased

excretion of base in the urine. This of course is accompanied by diuresis, marked dehydration results

On the other hand, the continued loss of pancreatic juice (p. 529) to the exterior causes an immediate depletion of base, plasma bicarbonate is reduced. In the adjustment of the acid-base balance the excess of acid radicles is excreted in the urine, this again entails a loss of water. Similarly, the ingestion of acid-producing salts causes a depletion of base, which is used for the neutralization and excretion of the acid radicles. Such salts therefore act as diuretics and dehydrating agents

In salt depletion the extracellular fluid is hypotonic, water is drawn into the cells, so that the volume of intracellular fluid is maintained, whereas the extracellular fluid (especially the interstitial) is reduced

Clinically the failure to ingest sufficient salt or the leaching of salt from the body of a seriously ill patient by glucose infusions is a not uncommon cause of salt depletion

3 *The injection of hypertonic solutions* (p. 37) into the blood stream. When a strong sugar or salt solution is injected, the temporary rise in the osmotic pressure of the blood causes a flow of fluid from the tissues into the vascular system until equilibrium is re-established. The blood volume is increased, but is soon returned to normal by the loss of the excess fluid into the tissues and its eventual excretion via the kidney and bowels. A net loss of body water results

Effects of dehydration

(a) *Loss of weight* due to the reduction in tissue water as well as to the actual breakdown of body substance which occurs in the effort to furnish water for the maintenance of physiological processes. Fat and carbohydrate stores are first drawn upon for this purpose and later, protein. (b) *Disturbances in acid-base balance*, usually toward the acid side. The diminished quantity of circulating fluid (loss of plasma water, anhydremia) and the consequent depression of oxidative processes in the tissues is held responsible for the excessive production of acid metabolites, e.g., lactic. The slowing of the renal circulation also leads to a reduced excretion of urine and the retention of acids (e.g., phosphoric) which under normal circumstances are eliminated. (c) *Rise in the non-protein nitrogen of the blood*. (d) *Rise in plasma protein concentration* and of chloride though there is no absolute increase

There is an absolute increase in blood sugar, especially when the stage of exhaustion approaches (e) *Rise in body temperature* as a result of the reduction in circulating fluid (see ch 4) (f) *Increased pulse rate and reduced cardiac output* (g) *Thirst* This occurs in water depletion but not in salt depletion Under normal circumstances thirst serves as a signal that the water stores of the body require replenishment Any fall in the water content of the tissues is reflected in the glandular activities especially of the salivary glands Secretion is suppressed, the mouth and throat become dry and the sensation of thirst is aroused In dehydration thirst is extreme and the mouth parched (h) *Dryness, wrinkling and looseness of skin* and a pinched expression to the features result from the loss of subcutaneous fat and of water from the deeper layers of the skin Other manifestations are, reduced intraocular tension and recession of the eyeball and, in infants, depression of the fontanelle (i) *Exhaustion and collapse*

WATER INTOXICATION

When an animal is given large quantities of water by stomach tube, especially if urinary secretion is reduced by the administration of pitressin, the tissues become "water-logged", serious symptoms ensue, e.g., depression of temperature, vomiting, convulsions and coma, which shortly end in death Similar effects also follow in man if large quantities of water are given to a patient with nephritic edema or, if in a subject of diabetes insipidus, pituitrin be administered while the water intake is maintained at the usual level (see ch 57) The manifestations of water intoxication are believed to be due to the dilution of electrolytes in the body fluids, and the damage caused thereby to the tissue cells ⁶

Adrenalectomy reduces the renal response to water drinking, and thus increases the susceptibility to water intoxication This susceptibility is reduced by the administration of desoxycorticosterone, of 17-hydroxy-11-dehydrocorticosterone, or of thyroid hormone A similar protective action is exerted by these principles upon normal animals Susceptibility of the latter to water intoxication is enhanced by the administration of the

⁶ Excessive concentration of electrolytes in the tissue fluid with consequent hypertonicity, as occurs in shipwrecked sailors if they drink sea water (which has a concentration in salts about three times that of serum) causes an equally deleterious effect upon the tissue cells This appears to be the cause of death

antidiuretic substance prepared from liver (p 77) and the discharge of this substance into the circulation may be a factor in the development of the condition, liberation of the antidiuretic hormone from the pituitary does not appear, however, to play any part in this connection

In water intoxication, the protein and chloride of the plasma are diminished and the extracellular water *decreased* The water retained in the body enters the cells of blood and tissues which become swollen It is not possible to account for the reduction in plasma chloride by increased renal excretion for both adrenalectomized and normal animals actually excrete less salt than usual, apparently the salt is diverted from extracellular to intracellular fluids

There is no danger from excess fluid being retained in the body through water drinking, for the sense of thirst and its appeasement nicely control the quantity ingested, but the artificial administration of inordinate amounts of water in the form of glucose solution, especially after surgical operations when there is some tendency toward anti-diuresis, may cause serious disturbances in water metabolism The effects of an excess of body water induced in this way can be corrected readily by the administration of hypertonic saline

ALTERATIONS IN BLOOD VOLUME

REDUCTION OF THE BLOOD VOLUME

This may result from

- (1) A loss of *whole blood* as in hemorrhage (p 27)
- (2) *Reduction in the total number of red cells*, as a result of increased destruction or diminished production (see anemia p 80)
- (3) Loss of *plasma* alone from the vessels as in extensive burns (p 305) or
- (4) *Loss of blood water* This is called *anhydremia* and is simply a part of a general dehydration and so results from the same causes as the latter ⁷

In the reduction of blood volume resulting from hemorrhage, the concentration of the blood in cells and protein is lowered, since a watery fluid is attracted into the vessels from the surrounding tissues

When the blood volume is lowered as a result

⁷ Variations in blood volume from time to time due to alterations in blood water may be detected from estimations of the hemoglobin or protein concentrations, or by means of the hematocrit

of a loss of plasma, the red cell concentration is increased (hemoconcentration) but the protein of the plasma is little altered

In anhydremia both the protein concentration of the plasma and the red cell count are raised. The concentration of plasma protein may increase by 50 per cent or more. The viscosity of the blood is therefore raised, the blood appears "syrupy" and flows sluggishly from an opened vein. If the anhydremia persists the red cell count and the protein concentration tend to fall again as a result of red cell and protein destruction. Then an estimation of the blood concentration may fail to give a true index of the extent of the blood volume reduction.

Exposure to cold causes a moderate loss of water from the blood to the tissues (chiefly skin muscle and probably liver), the total water content of the body remaining unaltered. It is not altogether clear by what means this movement of water is brought about. It is, however, an important factor in the regulation of body temperature (ch 54). The work of Barbour and others furnishes evidence of a nervous element in the mechanism. Animals in which the cord had been divided in the upper thoracic region when placed in a cold bath did not respond in the normal fashion. Concentration of the blood did not result and the temperature of the body fell to that of the environment. The control is exercised evidently through vasomotor nerves.

Barbour and Hamilton have shown that cold applied locally after section of the splanchnic nerves causes anhydremia as a result of the transudation of water into the skin of the cooled area. They attribute the migration of water to constriction of the cutaneous arterioles and consequent slowing of the capillary blood flow, which in turn, possibly through oxygen lack, increases the permeability of the capillary wall.⁸ It is probable therefore that exposure to cold induces blood concentration through both central and direct peripheral effects.

Posture The blood volume of the human subject after 30 minutes or so in the erect posture is some 15 per cent less than that in recumbency. A fluid of low protein concentration leaks from the vessels of the lower limbs into the extracapillary tissues, as a result apparently of the increased hydrostatic pressure in the capillaries of these parts (ch 35).

⁸ In this experiment, however, the cold was of such a degree that it may have directly damaged the capillary epithelium.

INCREASE IN BLOOD VOLUME

(1) *High temperatures* Two factors are concerned in the elevation of the blood volume which follows a rise in environmental temperature: (a) contraction of the spleen whereby whole blood is discharged into the general circulation (p 70) and (b) dilution of the blood, water being drawn from the tissues to augment the circulating fluid. This is a reversal of the mechanism described above as occurring at low temperatures. Sweating, and increased evaporation from the body surface, if over a prolonged period, will tend to counteract these effects, the blood volume may then show a decrease, or one effect may balance the other and no change occur.

(2) *Muscular exercise* At the beginning of exercise the blood volume is increased as a result of the discharge of blood from the spleen. Inasmuch as the splenic blood is relatively rich in red cells, the blood of the general circulation shows an increased cellular concentration. Later, as osmotically active substances (e.g., lactic acid) are formed in the contracting muscles, water is attracted from the vessels. The protein concentration of the plasma, and the red cell count increase. Sweating, when it ensues, tends to increase the degree of anhydremia by causing a reduction in the water content of the body as a whole. The early increase in blood volume which, in animals, results from the discharge of blood from the spleen is not seen, as a rule, in man. In dogs, muscular training causes an increase in blood volume and of the erythrocyte count which persists for about a month after the termination of the training period.

(3) *Emotional excitement* in animals and in man causes an increase in blood volume due to contraction of the spleen (p 70).

(4) *Pregnancy* Barcroft and his associates observed in sheep a pronounced increase in blood volume in the first and last thirds of pregnancy due to an increase in plasma. The corpuscular volume showed a relative decrease.

In pregnant women the plasma water as well as the interstitial fluid is increased, the latter by as much as 3 liters. There is an associated retention of sodium. Restriction or the administration of salt reduces or increases, respectively, the volume of extracellular fluid. The increase in plasma volume in pregnancy, combined with the lower protein concentration of the plasma which results from dilution and the consequent lowering in the plasma oncotic pressure, is responsible, in part at

least, for the edema of the lower limbs which commonly occurs in the pregnant state

(5) *Congestive heart failure* As a result of the retention of sodium and chloride (see p 38), the plasma and interstitial fluids undergo a pronounced increase

(6) *The administration of desoxycorticosterone* causes a retention of salt and, as a consequence, a rise in the volume of plasma and interstitial fluid

SUMMARY OF THE PATHOLOGICAL STATES ASSOCIATED WITH ALTERATIONS IN BLOOD VOLUME

Reduction

(a) *Hemorrhage* (loss of whole blood), (b) *burns* (loss of plasma), (c) *dehydration* (loss of water), (d) *pernicious anemia* (reduction in red cells, with a moderate increase in plasma), (e) *certain chronic anemias other than those of the pernicious type* In these the total volume of red cells is only slightly or moderately reduced and this is to a large extent compensated for by an increase in plasma above the normal standard The total blood volume is therefore, as a rule, not greatly below normal as calculated upon the basis of weight or of surface area (f) *Obesity* The blood volume per kilogram of body weight is much reduced but is normal when considered in relation to the body surface (g) *Myxedema* (reduction of both red cells and plasma but mainly of the former)

Increase

(a) *Polycythemia vera* (increase mainly of red cells but also of plasma), (b) *cirrhosis* of the liver (increase of plasma, see p 22), (c) *leukemia* (increase in white cells and plasma), (d) *splenomegaly* with anemia—Banti's disease (increase in plasma), (e) *hyperthyroidism* (equal increases both in red cells and plasma)

It should be pointed out that the proportions of red cells and plasma may vary from the normal though the total blood volume remains unaltered With regard to the blood volume and the proportion of cells to plasma there are therefore nine possible blood states Rowntree has introduced the following descriptive terminology A normal blood volume he terms *normovolemia* If the ratio of cells to plasma is normal as well, he calls the condition *simple normovolemia*, but a decrease or increase in the number of cells in relation to plasma is termed *oligocythemc* or *polycythemic normovolemia*

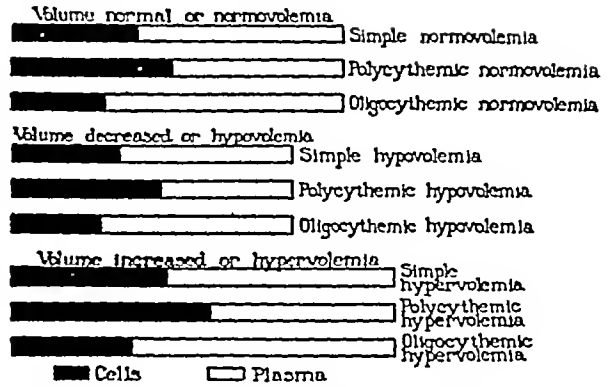


FIG 3 1 The nine possible combinations of whole blood, plasma and red cell volumes (after Rowntree)

respectively *Hypovolemia* and *hypervolemia* are corresponding terms for reduced and increased blood volumes, each of these is divisible into simple, polycythemic and oligocythemc forms (see fig 3 1)

THE EFFECTS OF HEMORRHAGE

When more than 30 per cent of the blood volume is lost rapidly, the body is usually unable to repair the loss unaided and, unless transfusion is resorted to, death results In a healthy man the loss of 30 per cent or less of his blood calls readjusting mechanisms into play which may bring the blood volume back to the normal level within a remarkably short time, 500 cc or so of blood drawn for transfusion purposes are said to be replaced within an hour or so The restoration of the blood to its previous concentration in erythrocytes, however, takes about seven weeks on the average This time may be shortened considerably by the administration of iron and a diet containing a liberal quantity of high quality protein Fowler and Barer found in a study of 200 blood donors that, after the removal of 550 cc of blood, the average fall in hemoglobin was 2 3 grams per 100 cc Regeneration of hemoglobin occurred at the rate of 0 049 gram per cent for men and 0 040 gram per cent for women per day

The protective mechanisms which automatically come into action after hemorrhage are several and may be divided into two groups—*immediate* or *early* and *delayed*

Immediate or early effects

If the loss of blood is large, especially if it is of sudden occurrence and from an artery, there is a prompt fall in blood pressure as a result simply of the reduction in circulating fluid (p 148) If not too great, the fall in pressure is salutary, since it

helps to prevent further bleeding. A moderate loss of blood, 10 per cent of the total amount, produces little or no drop in pressure. This is especially true if the blood is lost gradually and if it comes from a vein, i.e., beyond the peripheral resistance. Under these circumstances compensatory mechanisms easily maintain the pressure of blood at its normal height.

Clotting of the blood (ch. 12) which occurs within a few minutes serves to close the opening in the blood vessel. The blood tends to clot more rapidly than usual after a severe hemorrhage. The initial drop in pressure, when such occurs, aids the formation of the clot. Stanching of the bleeding is also furthered in the case of an artery by the retraction and contraction of the middle fibro-muscular coat of the vessel, as well as by the curling up and crenation of its endothelial lining. In the clotting process masses of agglutinated platelets appear, upon disintegration they liberate a vasoconstrictor material which is a material aid in hemostasis (ch. 12). These factors, together with the fall in blood pressure, may be sufficient to stanch the flow of blood from an artery as large even as the popliteal. In small vessels the opposed endothelial surfaces become sticky and coherent.

Increase in the heart rate. This is almost invariably an accompaniment of a severe hemorrhage and is one of the most valuable signs of concealed, i.e., internal, bleeding; this sign is not seen, as a rule, in moderate blood losses or in slow bleeding. It is brought about through carotid sinus and aortic reflexes (pp. 245 and 282) initiated by the fall in blood pressure. Reduction in blood flow through the vessels of the medulla with the consequent anoxemia of the cardiac centers may be an additional factor. Adrenaline liberation (ch. 59) may possibly play a part. The increase in heart rate might under the circumstances appear to be of advantage—an attempt on the part of the heart to increase the rate of blood flow and so compensate for the reduced volume of blood by having it carry an oxygen load from lungs to tissues more frequently. It seems clear, however, from the work of several investigators (p. 252) that simple increase in the cardiac rate does not increase the output of the heart when diastolic filling is inadequate.

Contraction of the spleen and the discharge into the circulation of a large quantity of blood rich in red cells (p. 70).

Increased respiration. The anoxia of the chemoreceptors of the carotid and aortic bodies caused

by the reduced blood flow (p. 405) is probably responsible for the increased rate and depth of breathing. When the blood loss is more profound, and consequently the oxygen want more urgent, long deeply drawn inspirations, and expirations of a sighing character ensue (air hunger), or periodic breathing of a Cheyne-Stokes type may develop (p. 413). Gasping respirations precede death.

Reduction in capacity of the vascular bed and redistribution of the blood. When the flow of blood has been stanching or considerably lessened by a complete or partial closure of the wound in the vessel, the blood pressure, if this had fallen, rises again. This is the result mainly of a readjustment of the capacity of the vascular system whereby it is made to conform more nearly to the lessened volume of blood. It is this reduction in the vascular capacity which prevents the initial fall in pressure when the loss of blood is gradual. It is effected by the reflex narrowing of innumerable small vessels (vasoconstriction) in regions such as the skin, mucous membranes, intestine and other parts not immediately essential to life. The vascular response is called into play by the underfilled state of the arteries and large veins feeding the heart (see vascular reflexes, chaps. 25 and 27).

The vasoconstriction is not confined to the small arteries and arterioles but extends to the capillaries, precapillaries and metarterioles (e.g., of muscles, mesentery and skin), which show increased reactivity to mechanical stimulation and to adrenaline (Zweifach and associates). This phase is associated with the appearance in the circulation of a vaso-excitor material (VEM) derived from the kidney. The discharge of adrenaline or nor-adrenaline probably also plays a part in the vasoconstriction. These measures whereby the blood remaining in the vascular system is confined to a smaller space are of the utmost importance; they enable the essential centers in the medulla to be supplied with blood under adequate pressure to sustain their vitality. Also a greater quantity of blood than would otherwise be possible is brought to the heart to supply its muscle, fill its cavities and maintain the circulation. The withdrawal of blood from the less important parts of the body is responsible, however, for some of the characteristic manifestations of hemorrhage, notably the pallor of the skin and mucous membranes, and the coldness of the body surface. The cerebral anemia causes sensations of giddiness or faintness, flashes of light or ringing in the ears (tinnitus).

The rise in blood pressure at this stage is con-

ductive to fresh bleeding. There is danger of the clot becoming dislodged.

In very severe hemorrhage, the hypernormal phase of capillary reactivity gives place to one of reduced capillary responses as the irreversible stage of hemorrhagic shock supervenes. The minute vessels tend toward dilation and become unresponsive, though the larger vessels still remain constricted. This phase is associated with the presence in the circulation of an hepatic vaso-depressor substance (pp 77 and 304).

Delayed effects

Replacement of the lost fluid. This, it has already been mentioned, commences almost upon the instant that the blood is lost (p 17), but takes a variable length of time, depending upon the extent of the blood loss, to become complete. Fluid is "drawn" from the tissues into the vessels and dilutes the blood. The corpuscular concentration is therefore reduced. The protein concentration of the tissue fluid is relatively low so that for a short time after hemorrhage the protein content of the plasma is markedly depressed. Very soon, however, the concentration of protein in the plasma shows a rise again as a result of the mobilization of protein stores. Calvin found, for example, that in dogs 50 per cent of the plasma protein removed by bleeding was restored within 4 hours. The extreme thirst which the subject of acute hemorrhage suffers is the call of the tissues for fluid and indicates that their own stores are being drawn into the under-filled vessels. The administration of water will therefore aid the body in recovering its water balance and replenishing the blood volume.

Replacement of the blood cells finally occurs through the increased activity of the blood-forming organs. This takes several days or weeks, the rapidity of the process depending to a large extent upon the nutrition and recuperative power of the individual and upon the diet (pp 27 and 75). While the repair process is in progress reticulated cells are found in increased numbers in the blood (p 16).

THE FACTORS GOVERNING THE INTERCHANGE OF FLUID BETWEEN THE TISSUES AND THE VESSELS

The physical factors which determine the flow of fluid from the tissues into the blood stream as well as in the reverse direction—from the vessels to tissue spaces—are the *osmotic* and *hydrostatic* pressures of the fluids in the two situations.

OSMOTIC PRESSURE

Osmotic pressure may be simply defined for our purpose here as the "attractive" or "drawing" force which sodium chloride, cane sugar and many other substances in solution exert upon the water molecules when water and a solution of one or other of these substances are separated by a membrane which allows the molecules of water to pass, but is quite or relatively impermeable to the molecules of the dissolved substance. Such a membrane is spoken of as "*semi-permeable*". An example will make this clear.

If an aqueous solution of cane sugar be placed in a vessel and a layer of water poured gently upon its surface, the two liquids will remain separate for a time. Gradually, however, sugar molecules will diffuse upwards and intermingle with the water molecules, while many of the latter will pass downwards into the sugar solution. The diffusion process, which is quite independent of gravity or convection currents, will continue slowly until the concentrations of the two types of molecules become equal throughout all parts of the liquid. If now instead of allowing free diffusion between the two solutions to take place, they be separated by a membrane which will permit the molecules of water to pass through it, but will offer a barrier to the migration of the dissolved substance, equal and free diffusion cannot occur. Since the water is able to pass into one compartment while the sugar molecules cannot pass out, the volume in this latter compartment obviously must increase, the pressure will rise. The pressure which is developed may amount to several atmospheres. This is the osmotic pressure, and it may be measured by the apparatus shown in fig 3 2 A. As the osmotic pressure within the inner chamber increases the mercury column is raised, that is, work is performed. In order to drive water through the membrane into the outer chamber

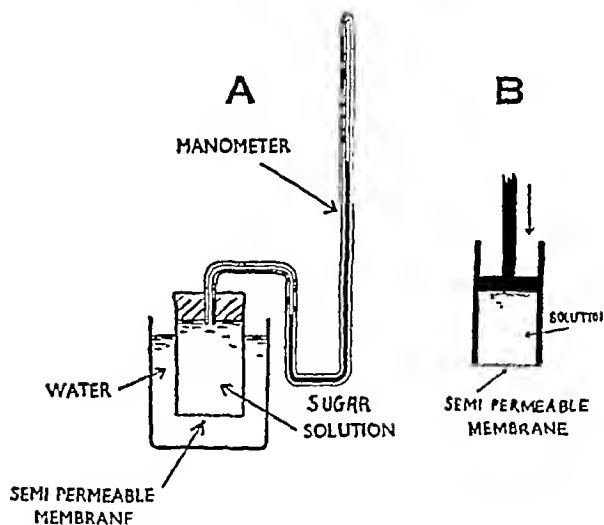


FIG 3 2 Description in text (after Parsons)

the osmotic force must be overcome. In fig. 32 B is illustrated a cylinder fitted with a piston and closed at its lower end by a semipermeable membrane. A force greater than the osmotic pressure developed within the cylinder would be required in order to filter water through the membrane. An instrument which measures osmotic pressure is known as an *osmometer*. An indirect method may be employed, e.g., measurement of the freezing point or the vapor pressure.

It is seen therefore from the foregoing that what was spoken of as an "attractive force" is simply the usual diffusion of water molecules into a solution of sugar which, owing to the intervention of a semipermeable membrane, cannot diffuse in the opposite direction. The actual pressure developed may be supposed to result from the bombardment of the walls of the containing vessel by the imprisoned molecules of the dissolved substance.⁹ The osmotic pressure of any solution will then necessarily depend upon the number of the bombarding molecules, i.e., upon their concentrations in the solutions. The pressure of a gas also depends upon the number of gas molecules per unit volume, and indeed it has been proved that the gas laws can be applied to the behavior of dilute solutions of cane sugar and other substances which exert osmotic pressure.

In other words, *substances in dilute solutions behave almost precisely as though they were in a gaseous state*. For instance, an amount of any gaseous substance equal to its molecular weight in grams (gram molecule) at standard temperature and pressure has a volume of 22.4 liters. That is, 32 grams of oxygen at 0°C and 760 mm Hg occupy the same volume (22.4 liters) as 2 grams of hydrogen under the same conditions. Also, according to Avogadro's hypothesis, equal volumes of gases at the same temperature and pressure contain equal numbers of molecules, i.e., their molecular concentrations are identical. The molecular weight of cane sugar is 342 grams. Therefore, according to Avogadro's law 342 grams of cane sugar if in the gaseous state would have a volume, at standard temperature and pressure, of 22.4 liters. Conversely, 342 grams of cane sugar if confined to a volume of 22.4 liters by dissolving it in that amount of water will, if the gas laws apply, have a pressure of 760 mm Hg at 0°C. Again, according to Boyle's law the pressure of a gas is inversely proportional to the volume at constant temperature. If therefore the same amount of cane sugar be dissolved in half the amount of water (11.2 liters) the osmotic pressure will be doubled. On the other hand, if the concentration of the molecules of sugar be reduced by dissolving them in double the quantity of water (44.8 liters) the osmotic pressure of the solution will be halved. Also, as in the case of a gas, the osmotic pressure of a substance in solution is proportional to the absolute temperature.

⁹ Several theories, of which this is the simplest, have been advanced to account for the phenomenon of osmotic pressure.

The gas laws apply with remarkable precision to dilute solutions, but in the case of concentrated ones, though there is a general agreement with the behavior of gases, other factors enter in which make the applications less perfect. Also substances generally classed as electrolytes (p. 126), sodium chloride for instance, exert osmotic pressures which are much higher than those of non-electrolytes, such as cane sugar or glucose. This apparent departure from the general principles is due to the fact that each molecule of the electrolyte is split into two parts (ions) and each of these acts as a separate particle and exerts its pressure effect. A sodium chloride solution, therefore, as a result of this dissociation has twice as many active particles as a cane sugar solution even though the number of molecules in a given volume (molecular concentration) of each be the same.

The semipermeability of the membrane, it must be emphasized, is an essential factor in the development of osmotic pressure. A solution may possess a very high osmotic pressure when separated from water by a membrane which does not allow the dissolved molecules to pass, but will exhibit no osmotic pressure when separated from water by a membrane perfectly permeable to the substance in solution. Certain membranes, for example the envelope of the red cell, are impermeable to potassium, sodium and calcium, but permeable to water. Others—such as the capillary wall—offer little hindrance to the passage of crystalloids. In the former case, osmotic pressure changes can be brought about within the cell by altering the concentration of sodium chloride in the plasma (p. 63). But in the latter case sodium chloride and other crystalloids, e.g., glucose, urea, bicarbonate, etc., though in the concentration in which they exist in plasma are capable when placed in an osmometer of developing a pressure of several atmospheres, exert a negligible effective osmotic pressure within the capillary since the membrane of the latter is freely permeable to them.¹⁰ The colloid osmotic pressure of the plasma is less than 1 per cent of the total osmotic pressure, i.e., of the osmotic pressure as measured in an osmometer.

Osmotic pressure is one of the fundamental forces underlying many physiological processes in both animal and plant life, e.g., the excretion of urine, the interchange of materials between the interiors of the blood cells or tissue cells and their surroundings, the flow of

¹⁰ It should be mentioned however that sudden changes in concentration of these substances (as by intravenous injection) will result in temporary disturbances of osmotic relationships while they are diffusing to re-establish equilibrium.

sap in plants, as well as the regulation of the blood volume. The fluids of the body contain various electrolytes and organic materials in solution. Semipermeable membranes of various types possessing different selective permeabilities, such as the cell wall, the vascular endothelium, the renal epithelium, and the membranes lining the serous cavities and the alimentary tract, are interposed between fluids capable of developing different osmotic pressures. The osmotic pressures are however not constant, but vary from time to time as a result of metabolic processes and the changes in concentration of various constituents of the intracellular and extracellular fluids incident thereto. Furthermore, the permeability of living membranes is not fixed and unalterable but is modified by several factors. Sodium and potassium chlorides, for example, increase membrane permeability to water, the chlorides of calcium and magnesium decrease it, permeability also probably is increased by a rise in temperature, and the eggs of various marine forms (*Echinoderm*, *Arbacia* etc.) become more permeable after fertilization. Narcotics, on the other hand, in non-toxic doses decrease the permeability of the cell to water.

ISOTONIC, HYPERTONIC AND HYPOTONIC SOLUTIONS

When two solutions are placed one on either side of a semipermeable membrane and the molecular concentrations of the dissolved substance are such that no osmotic pressure is developed, the solutions are said to be *isotonic*. That is, the pressures on the two sides of the membrane precisely balance one another. When one solution has a higher osmotic pressure than the other it is said to be *hypertonic*, the solution of lower osmotic pressure is termed *hypotonic*.

HYDROSTATIC PRESSURE

The other important factor in the interchange of fluid between the blood and the general body fluids is the hydrostatic pressure within the capillaries, i.e., the blood pressure, and its relation to that of the extravascular fluids.

THE RELATION OF OSMOTIC TO HYDROSTATIC PRESSURE IN THE INTERCHANGE OF FLUID ACROSS THE CAPILLARY MEMBRANE

It has already been mentioned that the substances in true solution in the plasma, such as glucose, inorganic salts, etc., exert little or no effective osmotic pressure within the capillaries. Their molecules are of such a size that they pass relatively freely through the "pores" of the mem-

brane. Obviously this must be so, otherwise essential nutritive materials could not reach the tissue cells and waste products could not enter the blood stream to be excreted. It is otherwise with the plasma proteins which, owing to the very large size of their molecules, cannot pass readily through the normal capillary wall. The latter is not, however, quite as impermeable to the plasma colloids as has been thought—they "leak" into the extravascular spaces—so, the osmotic pressure which they exert in the vessel is somewhat less than the value obtained in the laboratory. Since the albumin has the smallest molecule of the three plasma proteins, it escapes through the vessel in relatively greater amounts than the globulin and fibrinogen fractions.

The manner in which these two pressures—osmotic and hydrostatic—act in regulating the interchange of fluids between the tissues and the muscles may now be seen. The blood at the arterial end of a capillary has a pressure, let us say, of 30 mm Hg. This is a force driving the water and the dissolved crystalloids through the capillary membrane. But the hydrostatic pressure of the tissue fluid on the outer side of the membrane offsets, in part, that within. The pressure of fluid in the tissue spaces is difficult to determine but it is considerably less than that in the capillaries. It probably varies considerably in different regions, being low in those containing much loose areolar tissue. For purposes of illustration let it be assumed to be 10 mm Hg. The hydrostatic pressure, therefore, which is effective in forcing fluid out of the vessel (filtration pressure) is only the difference between the pressure within and that on the outside of the vessel, that is, 20 mm Hg. The osmotic (protein or oncotic) pressures of the plasma and tissue fluids must be taken into account, however. In the plasma it amounts to about 25 mm Hg. The tissue fluids have a lower protein content and consequently a lower osmotic pressure. The latter amounts to about 15 mm Hg. The difference, i.e., 10 mm in favor of the plasma will act as an attractive force to hold fluid within the vessels and so should be subtracted from the value of the effective hydrostatic pressure of 20 mm as calculated above. The net result of these opposing forces will be the filtration of fluid through the vessels under a pressure of $(20 - 10) = 10$ mm Hg.

The osmotic pressure of the plasma tends to rise as the blood flows through the capillary as a result of the passage of water outwards and the

consequent rise in the concentration of protein. That is, the force holding fluid within the vessel is increased. The hydrostatic pressure, on the other

Blood		Tissue fluid
Hydrostatic pressure 30 mm Hg	Capillary wall	Hydrostatic pressure 10 mm Hg
Effective hydrostatic pressure 20 mm Hg		
Osmotic pressure 25 mm Hg		Osmotic pressure 15 mm Hg
Effective osmotic pressure 10 mm Hg		
Driving force \rightarrow 10 mm. Hg (20 - 10)		

hand, falls gradually from the arterial to the venous end of the capillary. Near the arterial end, the blood pressure being greater than the osmotic pressure, a filtration of fluid with a low concentration of protein will result (see fig. 3.3). Near the venous end, the hydrostatic pressure falls below the osmotic pressure and a flow of dilute fluid (water and salts) from the tissue spaces into the blood takes place.¹¹ Transudation of a dilute plasma and reabsorption of a saline fluid, respectively, are continually going on in these two regions of the capillary bed.

Metabolic processes in the tissues bring about changes whereby larger molecules are being broken down into smaller ones, other molecules are removed or built up into larger ones. In this way alterations in molecular concentrations and in the diffusibility of the constituents of the tissue fluids

¹¹ It has been reported by McMaster and Rous that dye particles escape more readily from the venous end of the capillary. They conclude that the capillary permeability increases progressively from the arterial to the venous end. This implies that a steady fall in osmotic pressure would result as the venous end was approached since more protein would leak out. If this conclusion is correct, it is difficult to see how filtration and absorption could occur in different parts of the same capillary and it would be necessary to consider the possibility that filtration occurs in some capillaries and absorption in others.

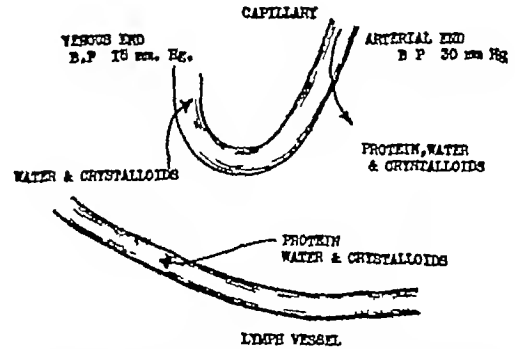


FIG. 3.3 Illustrating the relation of hydrostatic and osmotic pressures in the regulation of the interchange of fluid between the vessels and the tissue spaces

with consequent variations in osmotic relationships are occurring ceaselessly.

Under any circumstance in which the blood volume is increased or diminished, either the hydrostatic pressure or the osmotic pressure or both are altered, and it is through such changes that the blood fluid is restored automatically to its previous level. After hemorrhage, for example, the hydrostatic pressure is lowered in the capillary area but the osmotic pressure is unchanged. Fluid will therefore be absorbed from the tissue spaces. Moreover, constriction of the capillaries tends to reduce, dilatation to increase the capillary pressure. Therefore, the capillary constriction which follows a severe hemorrhage will slow the filtration rate at the arterial end of the capillary and increase absorption of interstitial fluid at the venous end. Again, when water is drawn from the blood, in consequence of excessive loss of fluid by the kidney, sweat glands or bowels, concentration of the plasma proteins will result. The increased osmotic pressure of the plasma will then hasten the rate of absorption from the tissues. The intravenous injection of large quantities of isotonic saline, on the other hand, will have the twofold effect of diluting the colloids and temporarily increasing the hydrostatic pressure. The excess fluid in consequence is rapidly eliminated from the blood stream into the tissues and later through the kidney and bowels.

CHAPTER 4

THE LYMPH AND TISSUE FLUIDS

STRUCTURE OF THE LYMPHATIC SYSTEM

The lymphatic system commences peripherally as a meshwork of delicate vessels (lymph capillaries) which drain the tissue spaces. By the confluence of small vessels larger ones are formed which, receiving tributaries along their course, gradually swell in size, and finally form the right lymphatic and thoracic ducts. These pour their lymph into the blood stream by way of the right and left subclavian veins, respectively. The system is a closed one, its vessels possessing complete walls formed of endothelial cells.¹ But their walls as compared with those of the blood capillaries are extremely permeable, a dye such as T-1824 (Evans blue), which escapes very slowly from the blood stream, passes rapidly from the lymphatic vessels. Small nodes (lymph nodes or glands) are interposed in the course of the larger lymph channels. These vessels, upon reaching the gland, break up into finer channels which, plunging into the node, open into the sinuses of its cortex. After passing through the gland the lymph is collected again on the other side by fine vessels which soon re-form into a few larger trunks. The glands are placed at strategic points along the lymph routes, e.g., the elbow and axilla, knee and groin in the case of the upper and lower limbs, and at points in the abdo-

¹ It is very widely though not universally believed that the lymphatic vessels have not an independent origin but are developed from the endothelium of the veins, appearing in the embryo as offshoots or "buds" from the internal jugular, abdominal and iliac veins (fig 4 1). The lymph vessels grow and extend by the proliferation of the endothelial cells composing the walls of these original sprouts. Such reservoirs of fluid as the subarachnoid spaces, the anterior chamber of the eye, the spaces of the internal ear (scala vestibuli and tympani) and the serous cavities (peritoneal and pleural) may be looked upon as belonging to the system of tissue spaces and as having developed from the dilatation and coalescence of smaller preexisting spaces in the mesenchyme. The fluid in the spaces of the central nervous system differs however from lymph and is drained, not by lymphatics, but by special absorbing structures—the arachnoid villi. Nor is the anterior chamber of the eye related to the lymphatics, but is drained through spaces in the pectinate ligament, and the aqueous humor differs in composition from lymph. Fluid is absorbed from the peritoneal cavity mainly via the blood capillaries. Solid particles (e.g., granules), on the other hand, are picked up by large phagocytic cells and conveyed into the lymph system, especially those of the diaphragm.

men, thorax and neck where several lymph vessels join. Lymph vessels are situated in skin, in subcutaneous tissue, in the fascial planes of muscles, in the linings of the respiratory, gastro-intestinal and genito-urinary tracts, and in the capsule and septa of the liver. Those in the intestinal villi are known as *lacteals*.

In the walls of the abdominal cavity the lymphatics are most abundant on the under surface of the diaphragm, where the greatest lymphatic absorption of colloidal material and minute particles takes place. The respiratory movements hasten absorption from the abdominal cavity, probably by varying rhythmically the intra-abdominal pressure. Absorption also takes place into lymphatics of the omentum. Phagocytes play an important role in absorption through both the diaphragm and omentum. Absorption through the parietal peritoneum is slight and is mainly through the blood capillaries, which absorb only crystalloid solutions. The lymphatic system of the heart consists of intercommunicating plexuses lying beneath the epicardium and endocardium, and within the myocardium. Lymphatics are also present in the areolar tissue underlying the peritoneum and pleurae. They are absent from the central nervous system. In the lung the lymphatics extend no further than the respiratory bronchioles, the alveoli being quite devoid of lymph capillaries. Very active proliferation of the lymph capillaries occurs in an inflamed region.

The skin is supplied richly with lymph capil-

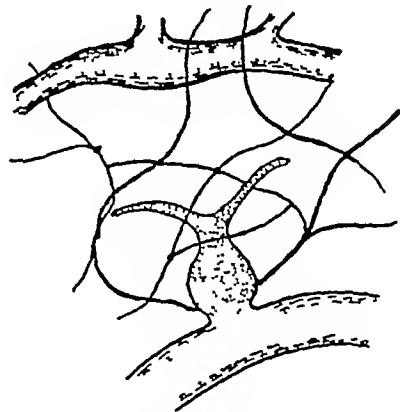


FIG 4 1 Showing a lymph vessel arising from a vein (redrawn from Sabin)

lanes The lymph vessels in the deeper layers of the skin are so abundant, according to McMaster, that the skin cannot be punctured anywhere without tearing them, and since the flow of lymph along these vessels is relatively rapid, foreign material injected into the skin soon reaches the regional lymph nodes An injection into the skin is, therefore, an injection into the lymphatic system The permeability of the lymph capillaries is increased by many agencies, e.g., sunlight, warmth, and by mechanical or chemical stimulation, histamine is particularly effective. Their walls may become so permeable that they can scarcely be considered as channels walled off from the surrounding tissue spaces

THE NODES AS DEFENSE BARRIERS

The lymph nodes must be looked upon as important structures for the defense of the blood against the invasion of bacteria or other injurious agents travelling by the lymph paths When an infection of a part—a finger for instance—lying distal to a gland occurs, the latter becomes inflamed as a result of the localization therein of some of the bacteria or their toxins carried in the lymph The gland swarms with motile cells (phagocytes) which attack and destroy the invading organisms In this way a barrier is raised against the passage of deleterious agents, particularly bacteria, into the blood stream In the case of the limbs at any rate it appears that no material can pass from the tissues to the blood stream via the lymph without filtering through the lymph nodes The effectiveness of the nodes as filters has been clearly demonstrated by Drinker and his associates The popliteal and iliac lymph nodes of dogs were perfused with solutions containing virulent streptococci (250,000,000 colonies per cc.) After perfusion lasting for over an hour the fluid collected from the thoracic duct was found to be sterile After the node itself has been attacked by the microorganisms it may then serve as a source from which the blood stream becomes infected Though highly efficient as filters for bacteria the nodes appear to offer but slight hindrance to the passage of viruses There is evidence that the lymph nodes contribute towards the body's defense in another way, namely, by the production of antibodies (p 95)

Barnes and Trueta have shown that bacteria pass from the tissues to the blood solely by the lymph stream, and that even toxins and venoms of

large molecular weight (over 20,000) are not carried into the blood if the lymph vessels have been blocked. These observations in part explain the success which has followed the immobilization of infected parts in plaster, this procedure might be expected to reduce to a minimum the lymph flow from the inflamed region and thus to confine the infective process

THE COMPOSITION OF LYMPH

The lymph of the small peripheral lymph vessels and the fluid of the tissue spaces are closely similar in composition, and both resemble the blood plasma They have, however, a protein content which is much lower than that of the blood fluid Drinker found the protein content to vary under different conditions from 0.3 to 4.0 per cent in mammals The higher figure, however, is unusual, concentrations between 0.5 and 0.7 per cent were found for human leg lymph

The proteins of the tissue fluid (and consequently of the lymph) are derived, as we have seen, from the plasma proteins and are identical with the latter, the three fractions alpha, beta and gamma being separable by electrophoresis But the proportion of albumin to globulin is greater (2 to 1) in lymph than in plasma, owing to the freer passage of the smaller albumin molecule The protein concentration is not constant, however, throughout the lymphatic system In the thoracic lymph it is as a rule considerably higher than in peripheral lymph, being about half that in plasma, and in lymph from the liver its concentration closely approaches that of plasma, being only about 16 per cent less The fibrinogen concentration of lymph, generally, is very low Lymph also contains prothrombin (p 113) It clots slowly Lymph contains large numbers of white cells, mostly lymphocytes, but relatively few red cells The number of leucocytes varies from 1000 to 20,000 per cubic millimeter in thoracic duct lymph of the dog and averages 550 per cubic millimeter in peripheral lymph In peripheral lymph there are from 300 to 13,000 erythrocytes per cubic millimeter The lymph flowing from the thoracic duct, since it comes largely from the intestine and liver, will vary in composition in accordance with the digestive processes Its protein content is under ordinary circumstances from 2 to 4.5 per cent Within 1 or 2 hours after a meal containing much fat the thoracic duct lymph appears milky The lacteals

TABLE 5
Chemical composition of peripheral lymph (cervical) and blood plasma from the dog
 (From Heim, 1933)

	PRO- TEIN (KJEL- DAHL)	NON- PRO- TEIN NITRO- GEN	UREA	URIC ACID	CREAT- ININE	SUGAR	AMINO ACIDS	CHLO- RIDES (AS NaCl)	PHOSPHORUS		CAL- CIUM
									Total	Inor- ganic	
	per cent	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc	mg per 100 cc
Plasma											
Average	6.18	32.6	21.7	Tr *	1.37	123.0	4.90	678	22.0	5.6	11.70
Range	5.54- 7.23	21.1- 46.0	17.9- 28.0		1.22- 1.54	112.0- 143.0		649- 721	18.3- 26.1	4.4- 6.9	10.85- 12.95
Lymph											
Average	3.32	34.8	23.5	Tr *	1.40	132.2	4.84	711	11.8	5.9	9.84
Range	1.38- 4.57	19.8- 45.4	19.8- 33.0		1.28- 1.49	107.0- 144.0		690- 730	10.2- 13.7	4.7- 7.3	8.93- 10.84
Number of animals	16	10	7	3	7	16	1	7	6	3	11

* Tr = trace

and the fine lymphatics of the mesentery, being loaded with globules of absorbed fat, are seen as glistening white streaks.

The composition of plasma and peripheral lymph (i.e., lymph in the subcutaneous vessels of the limbs or neck) are compared in table 5.

It will be noted that calcium and total phosphorus, which are in part bound with protein, are in lower concentration than in plasma, the other constituents with the exception of protein and amino acids are in higher concentration. The inequality in distribution of these electrolytes is explained satisfactorily by the Gibbs-Donnan Law.

THE FORMATION, PRESSURE AND FLOW OF LYMPH

After what has been said with regard to the forces concerned in the regulation of the fluid interchange between the capillaries and the tissues little need be added in explanation of lymph formation. The two processes are inter-related and similar in nature. Any condition which increases the outpouring of fluid from the capillaries into the tissues will tend to increase the flow of lymph. The lymph capillaries are much more permeable than the blood capillaries, and though the tissue spaces, i.e., the clefts between groups of tissue cells, are separated from the lymphatic system in an anatomical sense, the walls of the latter vessels are so permeable that they offer little impediment to the movement of either protein or crystalloids. It has been already stated that of the fluid which transudes from the blood at the arterial end of the

capillary much of the water is reabsorbed from the venous end. The protein, however, passes into the lymph. The lymph capillary is therefore the special channel whereby protein is returned (in a round about way) to the blood. It is also concerned with the absorption of other colloids, or of particulate matter which may be introduced into the tissue spaces. It is an interesting fact that in the frog, in which the blood capillaries are much more permeable to protein than those of mammals, the entire protein of the plasma passes from the blood and back again to blood via the lymph system, some 50 times in 24 hours (fig. 33).

Under ordinary conditions the pressure of the thoracic duct lymph is very low but if the duct is obstructed in the dog, a pressure of 15 cm. of water develops. The rate of flow along the human thoracic duct (as measured in cases of duct fistulae) is from 1 to 1.5 cc. per minute. Cain and his associates found an average flow of 0.46 cc. per minute in the thoracic duct, of which more than half, about 0.26 cc., was contributed by the lymph vessels of the liver.

The pressures in the peripheral lymph vessels during rest run from 0 to 6 cm. of water according to different observers using various species of animals. In inflammatory states and during activity the pressure in the larger lymph vessels of the part rises considerably above the resting value. Also when a lymph vessel becomes obstructed, the pressure on the distal side (i.e., toward the finer vessels) of the obstruction increases, the vessel

becomes distended on this side and collapsed on the proximal side. The lymph pressures decrease from the periphery toward the more central channels, and are higher on the distal side of a lymph node (mesenteric) than on the proximal side. In the lymph capillaries of the mouse's ear pressures up to 27 cm. of water were found by McMaster, the latter figure exceeds that usually found by most observers in the larger lymph trunks of other animals.

Obliteration of the lumen of the finer lymph vessels by a high pressure in the tissues is provided against by the fastening of the walls of the vessels by fibrillae to the surrounding tissue cells. When edema fluid collects and the tissue pressure rises, the fine lymphatics do not collapse, but, on the contrary, their lumina become wider (McMaster).

The mechanism governing the passage of tissue fluid into the lymph vessels is obscure, though the action of the pulse in the blood vessels of the part appears to play a part. McMaster and his asso-

ciates have demonstrated the importance of a pulsatile flow in the vessels of the perfused rabbit's ear, in the spread of vital dyes through the tissues and in the formation and flow of lymph.

CONDITIONS WHICH INCREASE THE LYMPH FLOW

(1) **INCREASE IN CAPILLARY PRESSURE AS A RESULT OF VENOUS OBSTRUCTION.** Landis and Gibbon found that in man filtration from the capillaries showed a definite increase when the venous pressure rose above 12 or 15 cm. of water. The rate of filtration from the capillaries was directly proportional to the increase in venous pressure (fig. 4.2). At a given venous pressure the filtration rate increased rapidly at first but gradually slowed and finally ceased. This falling off in the filtration rate is ascribed to the rise in extracapillary pressure, due to the fluid accumulation, which opposes the hydrostatic pressure within the capillary. The accumulation of extracellular fluid is therefore greater in regions which are loose in texture and where the skin is readily stretched. In persons with firm, resistant skin, edema, for the same reasons, is later in making its appearance and is less pronounced than in those with loose, flabby skins, as when weight has been lost rapidly.

Increased pressure in the veins of the portal area, as may be produced by obstructing the portal vein or the hepatic veins, causes increased filtration into the tissues of the abdominal viscera and a great increase in the volume of lymph flowing along the thoracic duct.

Increase in arterial pressure, on the other hand, does not increase filtration in animals until the pressure reaches around 300 mm. Hg.

(2) **INCREASED PERMEABILITY OF THE CAPILLARY WALL.** (a) *A rise in temperature increases capillary permeability, raises the filtration rate and the flow of lymph.*

(b) *Capillary poisons.* *Peptone* increases the flow of lymph from the thoracic duct probably as a result of its injurious effect upon the abdominal capillaries. The increased flow occurs after removal of the liver so injury to the hepatic vessels is not essential as was once believed for this action of peptone (Markowitz and Mann). Other substances which increase lymph flow in this way are extracts of strawberries, cray-fish, mussels and leeches, histamine and foreign proteins. Such materials are referred to by Heidenham, as *lymphagogues of the first class*. Heidenham held the view which is held no longer that the endothelium of the capillary

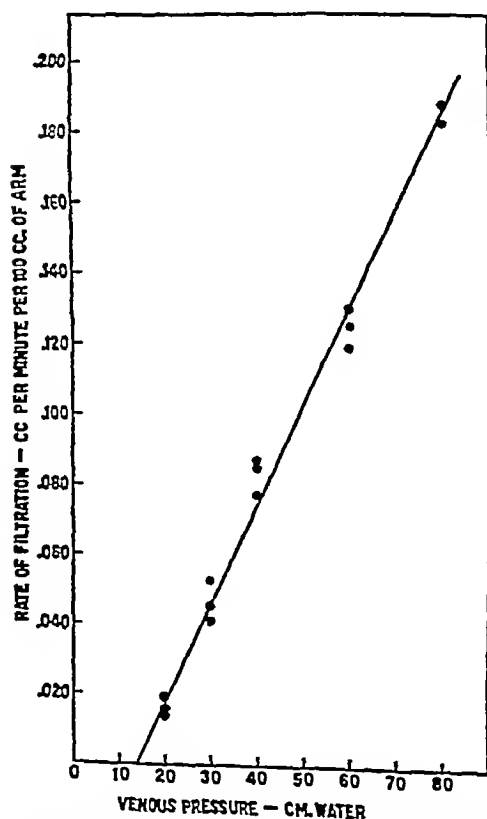


FIG. 4.2 Showing rates of filtration produced during 30 minutes by venous pressures between 20 and 80 cm. water (after Landis and Gibbon)

possessed a true secretory function and that it was stimulated to activity by lymphagogenic agents

(c) *Reduced oxygen supply* to the tissues (oxygen want), probably through damage to the capillary endothelium

(3) **HYPERTONIC SOLUTIONS** The intravenous injection of a concentrated solution of glucose, sodium sulphate or sodium chloride causes an increased flow of lymph from the thoracic duct in the following manner. These substances in concentrated solution, though they permeate the capillary wall, exert an osmotic pressure until equilibrium between the extravascular and intravascular fluids is re-established. Fluid is "attracted" from the tissue spaces, particularly of the muscles and subcutaneous tissues of the limbs, which in consequence show a fall in volume, the brain shrinks. The removal of fluid may actually extend to the fluids within the cells which undergo shrinkage, and general desiccation of the tissues results (p. 23). The blood volume for the time is greatly augmented, the excess fluid for the most part being accommodated in the capacious capillary and venous areas of the abdomen. The viscera—liver, kidneys, spleen and intestines—increase in volume, due to the distension of their vascular beds, and a great outpouring of fluid occurs, which swells the volume of lymph in the thoracic duct. In this way these substances produce a redistribution of fluid. Since the water-logged tissues eventually give up the extra fluid for excretion the body suffers a net loss.

The injection of isotonic saline will also increase the lymph flow since the plasma colloids are diluted thereby and the filtration through the capillary is increased. Substances which increase the flow of lymph in the manner described in this section are sometimes spoken of as *lymphagogues of the second class*.

The effect which hypertonic solutions have upon the movement of tissue fluids had extensive application during the first World War. Sir A. E. Wright introduced the practice of packing wounds with salt crystals or irrigating them with a hypertonic salt solution. This causes an outward flow of lymph and tissue fluid, which results in the mechanical removal of the bacteria and their toxins from the tissues bordering the wound.

(4) **INCREASED FUNCTIONAL ACTIVITY** When a gland or muscle enters into activity an increase in lymph flow occurs which starts a little after the commencement of the secretory or contractile response, but is nearly synchronous with the in-

creased metabolism resulting from the activity. The increased flow is ascribed to (a) formation of metabolites which increase the osmotic pressure of the tissue fluids and so "attract" more fluid from the vessels (b) vasodilation and increased capillary pressure.

During rest the flow along the lymph vessels of the muscles and subcutaneous tissues is slight, and the protein content of the lymph is high. During activity the protein concentrations fall since less transuded water undergoes reabsorption into the blood and more is carried away by the lymph channels. The contracting muscles exert a pumping effect upon the lymph, driving it along the vessels. The rhythmical movements of the intestinal wall exert a similar effect in promoting the flow in the fine lymphatics of the submucosa.

(5) **MASSAGE AND PASSIVE MOVEMENT** act to a certain extent like muscular activity. They augment the blood flow and capillary pressure and so increase lymph formation. The manipulations and movements of the muscles serve to propel the lymph along the lymphatic channels.

EDEMA (SEE ALSO P. 474)

Edema is the term applied to an excessive accumulation of fluids in the tissue spaces, and is due to a disturbance in the mechanisms of fluid interchange, which have been considered in the preceding pages.² Instead of there being a perfect balance struck between the inward and outward flow of fluid through the capillary membrane absorption is exceeded by transudation. The particular factor or factors of the mechanism that are

² When the subcutaneous tissues are involved these appear swollen, and leave the imprint of the thumb when it is pressed into the skin (pitting). Dropsy (hydrops) is an old-fashioned term which is applied to edema as defined above or to a free collection of fluid within one of the body cavities, e.g., the thorax or the abdomen. *Hydrothorax* is also applied to the former of these conditions and *ascites* to the latter. *Anasarca* is a more or less generalized edema involving the subcutaneous tissues.

It is important to remember that the term edema applies to a gross collection of extravascular fluid. The bulk of tissue fluid fluctuates widely in health, and in pathological states may be very considerably increased before the increase is evident clinically. When, for example, a normal person stands for a time the extravascular fluid of the legs increases, and the immersion of a limb in a hot bath hastens the rate at which fluid transudes from the vessels—the limb volume rises as a result, largely, of fluid accumulation in the tissues. Drury and Jones found that edema appeared when the increase in fluid increased the volume of the leg by 8 per cent.

disordered are not always clear, and a satisfactory explanation of all forms of edema cannot be given. But from previous discussions it is evident that the following factors will tend to increase the volume of interstitial fluid (1) Reduction in the osmotic pressure of the plasma, i.e., in protein concentration. Edema commences when the albumin fraction has fallen to between 2.5 and 3 per cent. (2) A general or a local rise in capillary blood pressure. (3) Increased permeability of the capillary membrane. (4) Increase in the filtering surface as when the capillaries dilate. (5) Obstruction of the lymph channels.

There is a tendency for edema to reach a certain degree and then to become stationary provided the conditions producing it remain constant, for as we have seen, when the tissue fluid pressure reaches a critical level its opposition to the force driving fluid from the vessels prevents further transudation.

Since edema is only a symptom of some primary condition it may have a variety of causes, according to the particular disease with which it is associated.

(1) *Cardiac edema* The edema of congestive heart failure has been generally attributed to an elevation of pressure in the venous and capillary systems combined with increased capillary permeability due to the anoxia resulting from slowing of the peripheral blood flow. The increase in venous pressure seen in congestive heart failure seemed to accord well with this view, for the edema is most pronounced in dependent parts, e.g., about the ankles in the standing position, where distension of the capillaries and the capillary pressure would be greatest. In some cases of congestive heart failure the plasma protein concentration is reduced, and this was looked upon as a contributory cause of the edema.

These explanations, viewed more critically, have proved inadequate. Warren and Stead have shown that in congestive heart failure an increase in the volume of extracellular fluid (plasma and interstitial) precedes any rise in venous pressure and that edema may commence with a normal venous pressure and plasma protein concentration. Moreover, an enormous increase in venous pressure follows ligation of the inferior vena cava, yet only a slight and temporary edema in the limbs occurs. Failure of the kidney to excrete the normal quantities of sodium chloride and water, which leads in turn to an increase in the volumes of interstitial fluid and plasma appears to be the fundamental

cause of the edema in congestive heart failure. A rise in the volume and pressure of interstitial fluid, according to Warren and Stead, precedes the enlargement of the plasma volume which is then followed by the elevation of venous pressure.

The renal disability, in the opinion of many, is due mainly to a reduced cardiac output rather than to a "back pressure" effect through the venous system with engorgement of the vessels of the kidney (see p 260).

When hypoproteinaemia exists it is not, as a rule, a prominent feature, and can act merely as a subsidiary factor in the production of this type of edema. Increased manufacture of plasma protein usually goes hand in hand with the rise in plasma volume so that the colloid concentration of the plasma remains, as a rule, unaltered and the total quantity of protein is increased.

It is questionable whether the capillaries are rendered more permeable as a result of anoxic damage to the endothelial wall. The existence of this factor in the production of cardiac edema has not been proved, but the observation that the edema fluid has a low concentration of protein argues strongly against it. Once a high venous pressure has developed, the filtration rate through the capillary membrane must necessarily be increased and the production of edema fluid facilitated. In addition to this the drainage of lymph from the edematous tissues appears to be at a standstill, for it has been observed by McMaster that though the lymph vessels are widely dilated they are filled with stagnant lymph. This phenomenon, which has not been accounted for and is not observed in other types of edema, must undoubtedly serve as an aggravating factor. (See table below.)

Summary of factors leading to the production of edema fluid in congestive heart failure

- A Main factors in the order of their occurrence
 - (a) Reduced cardiac output ("forward failure"), or increased renal venous pressure ("backward failure")
 - (b) Reduced renal blood flow
 - (c) Retention of sodium chloride and water
 - (d) Rise in volume of extracellular fluid
 - (e) Increase in plasma volume.
 - (f) Elevation of venous pressure.
 - (g) Failure of lymphatic drainage
 - B Subsidiary factors which may be present
 - (a) Low plasma protein, sometimes
 - (b) Anoxic damage to capillary wall
- Questionable.

The mechanism of sodium retention is still a controversial question

Merrill has investigated the factors involved in the retention of salt in congestive heart failure and has found that the renal plasma flow (ch 35) as measured by hippurate, was reduced by from 50 to 75 per cent or more of the normal, and the filtration rate, as measured by the inulin clearance method, by 25 to 50 per cent. The reabsorption of salt by the tubules was not increased, the retention of both sodium and chloride being due solely to the reduction in filtration rate, which was the result in turn of the reduced renal blood flow. This latter was diminished to a greater degree than was the cardiac output, which indicates a diversion of blood from the renal circulation.

There are other instances of cardiac edema which cannot be explained by a reduction in the glomerular filtration rate, this may be of normal value. Increased reabsorption of sodium by the renal tubules then appears to be the dominant factor. How this is brought about remains obscure, though it is suspected that some alteration in intracranial hemodynamics and fluid pressure is responsible. Sodium retention, for example, is less in the recumbent than in the sitting position, but raising the volume of the extracellular fluid within the cranium, by compression of the veins of the neck, increases the excretion of salt in the sitting position comparable with that in recumbency (Harrison). The rise in pressure in the renal veins in the upright position also tends toward a reduction in salt excretion. How reduction in intracranial extracellular fluid volume enhances the tubular reabsorption of sodium is quite unknown. The possibility has been suggested that it is of hormonal nature, through the suppression of the antidiuretic principle, or to increased activity of the adrenotrophic hormone upon the adrenal cortex. Increased excretion of adrenal steroids in cases of congestive heart failure have been reported.

As mentioned elsewhere, fluid in order to be held in the tissues must be isotonic, salt retention is, therefore, followed inevitably by an increase in body water. It is clear then that the volume of extracellular fluid can be reduced by restriction of the salt intake, or by the administration of diuretics which remove both water and salt. Also, para-

as by administering dig suitable drug

(2) *Mechanical obstruction*—main veins leading from new growth, fibrous tissue, liver, thrombosis, etc., fluid occurs. This is due to capillary pressure but capillary wall is also in impaired blood supply, growth, probably to the stances as well. In the cirrhosis, in which ascites is an additional factor in the liver of an antidiuretic

(3) *Edema due to glomerulonephritis* and due to the reduction in plasma proteins and so the blood (see p 474). If the capillary wall is damaged and protein escapes in excess into the tissues, so the edema is due to protein concentration.

(4) *Inflammatory edema*—factors combine to produce the edema. Increased permeability due to dilatation of the capillary wall of the blood stream (p 474) and obstruction of lymphatics for a variable time. Inflammatory areas are obstructed, capillary walls are also seriously injured. Toxin or other injurious substances with a high protein content are present. The edema is localized to the area surrounding the injured tissue, e.g., aspirin, morphine produce the edema of simple

(5) *Giant edema*—This is a form of edema which comes in the hands, face, external genitalia, runs in families. Little is known of its production. It is liberated at the site of its immediate cause (ch 28). It is frequently a foreign protein which apparently gains access to the tissues.

anaphylactoid nature and constitute one type of allergy. This type, also termed *agioneurotic* edema, is allied to the very localized edemas which constitute the condition known as urticaria and which as suggested by Lewis are due to the liberation of a histamine like substance in the skin.

(6) *Edema due to malnutrition or to toxic substances* Edema may occur in the anemias or in conditions in which the general nutrition of the body suffers. When the diet is deficient in vitamins, or there is too little fat or protein in the diet edema may occur, as in beri beri, scurvy, "war edema" or in the faulty nutrition of infants. In animals edematous conditions have actually been induced by general underfeeding, or by a diet deficient in fat and in fat soluble vitamins, or by one deficient in protein alone. The factors responsible for the increased transudation in these cases are not always clear, but in others there is a marked lowering of plasma protein which alone is sufficient to account for the edema. In many instances, on the other hand, hypoproteinemia does not occur. In such cases the lack of some essential amino-acid may be the determining factor. A simple explanation has been offered by Henschel and his colleagues, namely, that owing to the loss of tissue on the famine diet the extracellular fluids (plasma and interstitial fluid) show an apparent increase that is, though not showing an absolute increase, they are greater than normal in relation to body weight. Henschel and his colleagues in experiments upon a group of normal young men on a semi starvation diet found that, while the *absolute* volume of extracellular fluid remained fairly constant at the value observed before the subjects were put upon the experimental diet, it increased gradually to about 40 per cent above normal when considered in relation to body weight, edema appeared when the relative increase in volume reached from 8 to 10 per cent above normal, which is approximately the same as the increase in volume of interstitial fluid at which other types of edema appear (see footnote, p. 37). Increased capillary permeability due to impaired nutrition of the vascular walls was the main cause of the starvation edema in the Netherlands during the last war. In some cases of nutritional edema, excretion of an antidiuretic substance has been reported.

Certain chemical substances such as arsenic, salts of heavy metals, and the toxins of certain infectious diseases, such as diphtheria, acute nephritis, etc., are known to act as capillary poisons and apparently cause edema in this way. An interesting type of a toxic edema is that which may be produced in animals by the injection of hematinophyrin (p. 60). This substance appears to sensitize the tissues towards light rays, and the edema occurs only after exposure. Histamine causes local edema at the point of injection by inducing capillary dilatation and increased permeability of the membrane.

(7) *Edema due to lymphatic obstruction* Obstruction to the outflow of lymph from the tissue spaces may cause pronounced edema, even though the venous channels and the capillary vessels are unaffected. Edema of this nature is readily produced in frogs by compression of the lymph channels alone. It is more difficult to produce edema in this way in higher animals but if the obstruction is complete edema occurs in them also. Edema of this nature is seen in infections with the filarial parasite which finds its way into the lymph vessels of the limbs and blood—their lumina with the production of the condition known as elephantiasis. The pleural cavities depend for the absorption of a fluid upon the lymph channel, and accumulation of fluid may occur here as a result of lymphatic obstruction. The edema associated with carcinoma is due chiefly to the filling of the lymphatic channel with cords of cancer cells as well as to venous obstruction caused by the pressure of the growth. The "milk leg" of the periparturient is in part due to lymphatic obstruction. The tissue fluid in these types of edema has a relatively high concentration of protein.

(8) *Heat edema* The effect of heat upon capillary permeability has been mentioned (p. 35). Excessive heat may actually lead to edema in man. It occurs in the tropics and occasionally in so-called temperate zones during an intense heat wave. Increase in blood volume, enlargement of the filtering surface as a result of the opening up of fresh capillaries and the rise in capillary pressure incident to the dilatation of capillaries previously patent, are also factors in the production of this type of edema.

CHAPTER 5

TRANSFUSION THE BLOOD GROUPS

The materials employed for restoring the blood volume to normal are, (1) *Whole blood*, (2) *plasma or serum*, (3) *solutions of colloids*, e.g., gum acacia, isinglass etc., (4) *solutions of crystalloids*, e.g., saline or glucose solutions

(1) WHOLE BLOOD¹

Theoretically, whole human blood is, of course, the ideal transfusion fluid. The improvements and simplification of technique in recent years and the advance in knowledge of blood incompatibilities have made blood transfusion immeasurably safer, and rendered it available under circumstances which hitherto would have been insuperable. It is used not only in cases of emergency such as severe hemorrhage or wound shock, but in several other conditions. Below are listed some of the conditions in which it is employed

Hemorrhage

Anemias, especially, aplastic anemia and hemolytic disease of the newborn, agranulocytosis and hemorrhagic diseases, e.g., hemophilia and purpura hemorrhagica

Shock (wounds, burns)

Malnutrition in infants, marasmus, acute intoxications

Septic conditions, septicemias

The red cells of the transfused blood survive and carry out their functions for several weeks (around 80 days on the average) after their injection. On this account whole blood is greatly superior to any other transfusion fluid in any condition in which the respiratory area of the blood has been greatly reduced, e.g., very severe hemorrhage or hemor-

rhage in an anemic person, CO poisoning, etc. It is also more effective in traumatic shock than plasma, serum or other blood substitutes. The use of blood as a transfusion fluid, nevertheless, is hedged about by hazards both to the recipient and to some extent to the donor. On this account it is a suitable method only when adequate facilities for guarding against these dangers are available, otherwise some blood substitute will have to be resorted to. The safeguards which must be taken are

(a) The donor must be healthy in every way. Several cases have been reported of disease having been transmitted by transfused blood. Syphilis, malaria and acute diseases have been reported to have followed blood transfusion. In some instances the transmission in the transfused blood of a particular foreign protein of dietary origin, to which the recipient but not the donor is sensitive, has caused an allergic reaction.

A very serious hazard today is the transmission in the donor's blood of a dangerous type of infectious hepatitis. This disease, when contracted through blood transfusion, is called *homologous serum jaundice* or *hepatitis*. It closely resembles and may be due to the same agent, apparently a filtrable virus, which is responsible for the infectious hepatitis, seen in both sporadic and epidemic forms.

(b) A too rapid transfusion of blood is dangerous, especially in children or undersized persons, for the sudden increase in circulating fluid may cause serious embarrassment to the right side of the heart. In an adult, the usual transfusion rate is from 100 to 200 cc per hour, which causes little cardiovascular effect, even though the blood volume is considerably increased. At high rates of transfusion the venous pressure may rise unduly, the cardiac output be increased sharply, and the heart be called upon for the performance of an excessive amount of work. Rapid infusions of fluid into animals causes death from cardiac failure, preceded by an inordinate rise in venous pressure and pulmonary edema. The total quantity injected varies according to circumstances and the size of the patient, from about 500 to 1500 cc or more.

¹ Within recent years stored blood (or plasma) is being used to an ever increasing extent. Blood collected from the dead has been employed in Russia, but cadaver blood for obvious reasons has not found general favor. Placental blood or blood removed by venesection from cases of congestive heart failure has also been used, but the blood of healthy donors is preferable. The blood is preserved at a temperature of around 1°C after dilution with a citrate-dextrose mixture in the proportions of 5 parts of blood, 1 part of 3.2% citrate solution and 6.5 parts of 5.4% dextrose solution. Kept in this way blood remains suitable for transfusion for about twenty days. These so-called *blood banks* have the advantage that a quantity of blood already prepared for transfusion can be obtained at a moment's notice. But there is also the great disadvantage that any blood not used within ten days or so must be discarded.

(c) One of the unpleasant and disturbing but rarely dangerous effects of transfusion is the rise in temperature which follows a short time after injection, unless great precaution is exercised in preparing the blood and apparatus. This pyrogenic action which has been a more prominent feature of transfusions with materials other than whole blood is due to substances (*pyrogens*) of unknown nature, but probably protein, formed in the distilled water used in the preparation of the blood substitute, in the material itself or contaminating the transfusion apparatus. They are thought to be in most instances of bacterial or virus origin. Therefore, though every possible care has been taken in their preparation, all artificial transfusion materials must be submitted to a reliable test upon rabbits for pyrogenic action before they can be accepted for clinical use.

(d) An ever present potential danger in the use of whole blood is incompatibility. The donor's blood must always be tested for its compatibility with the blood of the recipient. Normal plasma contains substances which have the power to cause the clumping together (agglutination) and subsequent disintegration (hemolysis) of the foreign corpuscles of another species. Agglutination also results when blood of two persons belonging to certain blood groups are mixed. The

bloods are then said to be incompatible and transfusion under such circumstances will lead to very serious if not fatal results (p. 44).

THE BLOOD GROUPS

As a result of the work of several investigators in the early years of this century, notably that of Landsteiner, it has been established that the blood of any person falls into one or other of four well defined groups, according to its agglutinating reactions. These groups were originally designated by the Roman numerals I, II, III and IV, respectively, but are now referred to by the letters O, A, B and AB (see p. 44).² The groups exhibit the reactions shown in table 6.

About 45 per cent of all individuals of European stock belong to Group O and 42 per cent to Group A. Ten per cent and 3 per cent belong to Groups B and AB, respectively.³

Consultation with the table will show that the serum of Group AB (vertical row on extreme right) is compatible with the corpuscles of all the other groups, i.e., no agglutination of the corpuscles of any donor should occur when the recipient belongs to this group. It will be noted however, that the cells of Group AB (lowest horizontal row) are agglutinated by the sera of all the other groups when tested outside the body. It might therefore be thought that this reaction would occur in the blood of the recipient, but as a matter of fact, when a patient belonging to Group AB is transfused with the blood of any other group, agglutination of his (patient's) corpuscles does not usually occur. The reason for this is not clear, unless it is that the serum of the injected blood is so highly diluted by the patient's serum. It is said however, that if the donor's serum be diluted to the same degree as occurs when it is transfused and is then mixed with the cells of group AB outside the body, agglutination does occur. Whatever the explanation, the fact remains that the reaction of the

TABLE 6

Landsteiner's classification of blood groups

CORPUSCLES	SERUM			
	O	A	B	AB
O	—	—	—	—
A	+	—	+	—
B	+	+	—	—
AB	+	+	+	—

+ means agglutination, — means no agglutination.

Group O The serum which agglutinates the corpuscles of the other three groups. The corpuscles of this group are not agglutinated by any serum.

Group A The serum agglutinates the corpuscles of Groups B and AB, but not those of Groups O and A. The corpuscles are agglutinated by the serum of Groups O or B, but not by that of Groups A and AB.

Group B The serum agglutinates the cells of Groups A and AB, but not those of Groups O and B. The corpuscles are agglutinated by the serum of Groups O or A, but not by that of Groups B and AB.

Group AB The serum of this group does not agglutinate any corpuscles. The corpuscles are agglutinated by the serum of all other groups.

² The terminology now adopted obviates the confusion caused by two classifications which were in current usage, one introduced by Jansky and followed mainly in America, the other by Moss, used in England. In both classifications the groups were designated by Roman numerals I-IV, but group I of the Moss classification corresponded to Jansky's group IV and his group IV to Jansky's group I. Groups II and III were the same in both.

³ These are general figures, they are drawn from a number of surveys by several investigators. The figures for other races have a wide range from 22 to 100 per cent for group O, 0 to 55 per cent for group A, 0 to 36 per cent for group B and 0 to 18 per cent for group AB.

donor's corpuscles to the serum of the recipient is the most important factor to consider, and that the agglutinating property of the donor's serum is not generally evident. On this account members of Group AB are sometimes spoken of as "universal recipients"

It will be seen from the table that the corpuscles of a person of Group O (uppermost horizontal row) are not agglutinated by any serum, though his undiluted serum will agglutinate cells of all other groups (first vertical row on left of table). Members of this group are therefore called "universal donors" since agglutination does not, *as a rule*, occur when this blood is transfused into a member of any of the other groups.

The terms "universal donor" and "universal recipient" though in general use are dangerously misleading. Severe and even fatal reactions may occasionally result from the transfusion of blood of group O into a subject of one of the other groups. The untoward reaction, when it occurs, may be due to an unusually high agglutinin titer of the donor serum, which is then capable of causing agglutination of the patient's cells, to subgroup incompatibility or to the Rh factor (see pp 45 and 46).

Similarly, it cannot be taken for granted that a subject of group AB can be transfused with impunity with the blood from any of the other groups. Such procedures are especially hazardous in the case of children, probably for the reason that they are transfused with relatively larger quantities of blood than are adults, the donor's blood, in consequence, is in relatively high concentration in the recipient's blood stream. Therefore, the blood of the donor should always be matched directly with the patient's blood (p 44).

Determination of the group to which a particular blood sample belongs

In order to determine the particular group to which a given sample of blood belongs, it is not necessary to have sera of each group, but only of Groups A and B. Sealed tubes containing high titer sera of these two groups are kept on hand for testing purposes. Reference to tables 6 and 7 and to figure 5 1 will show the reason for this. If a diluted specimen of the unknown blood is agglutinated neither by Group A nor Group B serum, it must belong to Group O. If it reacts to Group B serum but not to Group A, it must belong to Group A, and similarly, if it reacts to Group A but not to

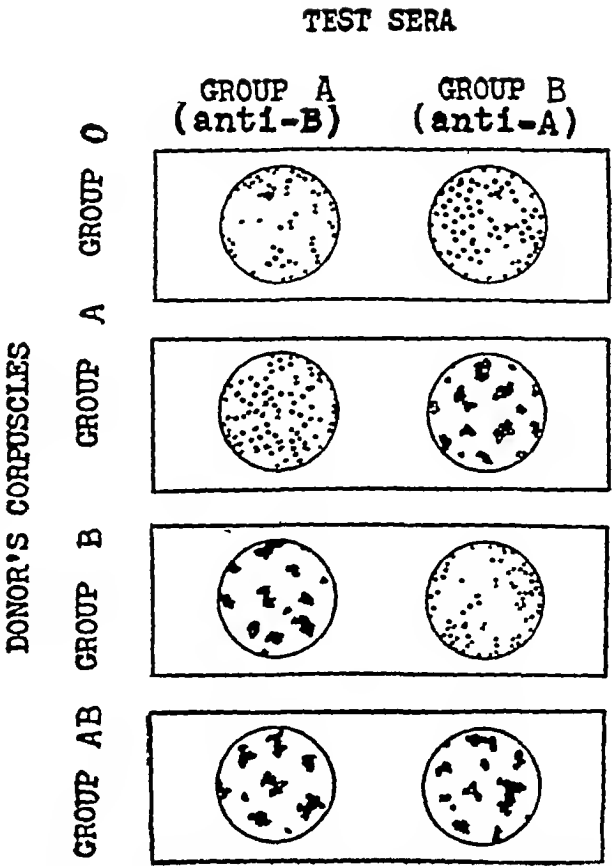


FIG 5 1 Showing the effects of the sera of groups A and B upon the corpuscles of the several blood groups

Group B, it belongs to Group B. If it is agglutinated by both test sera, it belongs to Group AB. To perform the test a drop each of Group A and Group B sera are placed side by side upon a white opal glass tile which has been slightly warmed. A specimen of blood from the person whose blood group is being determined is centrifuged and the cells mixed with a 0.9 per cent solution of sodium chloride to make a 2 to 5 dilution. A drop of the suspension is added to each sample of test serum.

Agglutination, if it occurs, is usually visible under the microscope within a few minutes or may be seen with the naked eye, the clumped corpuscles

TABLE 7
Reaction of red cells of unknown

UNKNOWN BLOOD	GROUP A SERUM	GROUP B SERUM	GROUP TO WHICH SAMPLE IS ASSIGNED
X	-	-	O
X	-	+	A
X	+	-	B
X	+	+	AB

+ = agglutination, - = no agglutination.

appearing like grains of cayenne pepper upon a clear fluid. The reaction, however, may be delayed for as long as 20 minutes. In the absence of agglutination the fluid remains uniformly pink. Lists of donors, classified in respect to the groups to which they belong, are usually kept in large clinics so that the appropriate blood can be obtained quickly in an emergency. Cross or direct matching should also be carried out as an additional check. A diluted specimen of the recipient's blood is mixed with a sample of the donor's serum, and a similar sample of the donor's blood with the patient's serum. The mixtures are observed microscopically for agglutination.

The effects of the transfusion of incompatible blood

The most urgent symptoms develop. Among these are tingling pains shooting through the body, severe lumbar pain, precordial distress, cyanosis, rapid thready pulse and other manifestations of severe collapse which frequently end fatally. Symptoms are due, in part at least, to the mechanical effect of the agglutinated corpuscles, which cause the blockage of small vessels in vital regions. Later, hemolysis of the agglutinated cells occurs. Hemoglobinuria, followed by renal failure with anuria, commonly occurs. The suppression of urine has been thought to be due to tubular damage caused by the excreted hemoglobin precipitated in the tubular lumen by the acid urine. It was, therefore, considered wise in such cases to alkalinize the urine. But experimental evidence is against this supposed role of the liberated hemoglobin, for Yuile and associates have shown that large amounts of hemoglobin can be excreted by dogs (after a transfusion with incompatible dog blood) without renal damage, and have suggested that anuria occurs in such transfusion accidents only when there has been some previous tubular damage or when coincident renal injury has been caused by some condition associated with that for which the transfusion had been undertaken.

TABLE 8

Showing the presence in the various groups of agglutinogens and agglutinins

GROUP	CELLS CONTAIN ISO-AGGLUTINOGENS	SERUM CONTAINS ISO-AGGLUTININS
O	O (neither)	α and β (anti A and anti-B)
A	A	β (anti-B)
B	B	α (anti A)
AB	A and B	<i>o</i> (neither)

Theory of blood grouping

Landsteiner postulated the existence of two specific substances in serum which he called *iso-agglutinins*, and two substances in the corpuscles which he termed *iso-agglutinogens*. The former may be represented by the Greek letters α and β , the latter by the capitals A and B. A given serum might contain one, both or neither iso-agglutinin. Similarly the corpuscles may contain one, both or neither iso-agglutininogen (table 8). In order for agglutination to occur when two bloods are mixed α must be present with A ($A\alpha$) or β with B ($B\beta$). Obviously, no such combination exists naturally in any blood, for auto-agglutination would result.

The incompatibility of two types of blood is dependent essentially on an immunity or antigen-antibody reaction comparable to the agglutination of bacteria by an antibody contained in immune serum. In the reaction of two types of blood the iso-agglutininogen is the antigen and the iso agglutinin the antibody (anti-A or anti-B). The antigen, however, is not introduced from without and no process of immunization is required for the production of the antibody, each is an inherent constituent of a certain type of blood.

If *o* be used to represent the absence of the two iso-agglutinins and O the absence of the two iso-agglutinogens then the four groups may be designated as follows:

Group O. The corpuscles have no iso-agglutininogen (O), the serum has α and β iso-agglutinins. The full designation of this group is, therefore, group $O\alpha\beta$. Its corpuscles obviously cannot be agglutinated by any serum but its serum will agglutinate any corpuscles containing A or B, i.e., the corpuscles of any of the other groups.

Group A. Corpuscles have A, serum has β . The full serological characters of this group would be represented by letters $A\beta$. Its corpuscles are agglutinated by any serum containing α , its serum agglutinates corpuscles containing B.

Group B. Corpuscles have B and serum has α . The complete designation of this group is $B\alpha$. The corpuscles are agglutinated by a serum containing β and the serum agglutinates corpuscles containing A.

Group AB. Corpuscles have A and B but serum has no iso-agglutinins. The group is fully represented by the letters ABo . Obviously the corpuscles are agglutinated by any of the other sera, but the serum of this group will not agglutinate the corpuscles of any other group.

It is owing to the confusion which has arisen from the existence of two classifications, Jansky and Moss, that this system has been adopted whereby the blood groups

are referred to by the letters indicating their serological characteristics, namely $O\alpha\beta$, $AB\beta$, etc., or simply by letters showing the characteristics of the corpuscles, that is, O, A, B and AB

The *titer* of a particular sample of serum, i.e. the agglutinin potency, is determined by titrating graded dilutions ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, $\frac{1}{32}$, and so on) with a suspension of sensitive red cells containing the corresponding agglutino-gen. The titer is expressed as the reciprocal of the highest dilution at which agglutination is observed. Thus, if the highest dilution at which the serum is effective is 1 to 32, the titer is 32

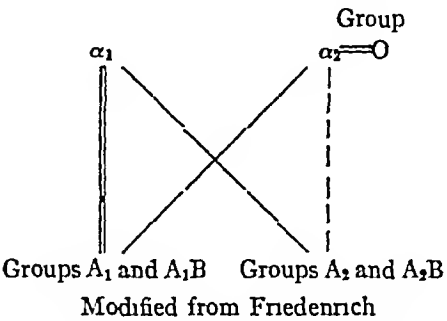
A, B and AB specific substances of Witebsky Substances of a carbohydrate-like nature, having A, B, or AB group activity, are present in high concentration in certain tissues and secretions of the body, e.g. saliva and gastric juice. They are also present in amniotic fluid but have not been demonstrated in the red cells. These substances, when injected in minute amounts into the blood stream, induce iso-immunization, increasing the titers of the corresponding agglutinins several fold. When added to blood *in vitro*, they reduce the activity of the corresponding agglutinin. A and AB specific substances are most readily available, B substance is difficult to obtain in the pure state. These substances have been employed to reduce the agglutinin titer of O group ("universal donor") blood which, as mentioned on page 43, may possess a dangerously high titer of anti-A and anti-B agglutinins. The addition of A and AB specific substances to such blood reduces the agglutinins by from 75 to 95 per cent. The specific substances have other practical applications. In sera used in testing for the Rh-factor, the actions of isoagglutinins anti-A and anti-B must first be annulled. This can be done by the addition of the specific substances to the test serum. Also, these substances are used in small amounts to produce anti-A and anti-B test sera of high titer for, as mentioned above, they stimulate the production of these iso-agglutinins. The reader is referred to papers by Witebsky and his associates for full accounts of these substances

Subgroups A₁ and A₂ Some years after the four blood groups were discovered and described by Landsteiner, it was found (Dungern and Herszfeld, 1911) that group A was made up of groups A₁ and A₂. That is to say the agglutino-gen A is of two kinds—A₁ and A₂. Since either of these might be inherited with B, there are also two subgroups of AB, namely A₁B and A₂B. About 80 per cent of persons belonging to group A are included in subgroup A₁ and about 20 per cent in subgroup A₂. In rare instances, the agglutinin α_2 is present in the plasmas of subgroups A₁ and A₁B and agglutinin α_1 in the plasmas of subgroups A₂ and A₂B. The serum of the main group O ("universal donor") may also, though very rarely, contain α_1 agglu-

tinin. See table below. Should the plasma of a donor contain α_1 agglutinin (anti-A₁) in high titer and the recipient belong to subgroup A₁ or A₁B a reaction will occur.

α_1 agglutinin reacts strongly with the cells of subgroups A₁ and A₁B and less strongly with the cells of subgroups A₂ and A₂B. α_2 serum agglutinates A₁ and A₁B cells, but reacts very weakly, or not at all, with cells of subgroups A₂ and A₂B. α_2 also reacts strongly with cells of group O. For this reason a fatal reaction may result from the transfusion of blood from a so-called universal donor.

These relationships are shown in the accompanying scheme and table.



The double lines represent a strong reaction, the single lines a weaker one, and the broken line a very weak reaction or none at all.

BLOOD GROUPS	AGGLUTINOGENS (IN CELLS)	AGGLUTININS (IN PLASMA)
O	None	α and β (rarely α_1)
A {	A ₁	β (rarely α_2)
	A ₂	β (rarely α_1)
B	B	α
AB {	A ₁ and B	None (rarely α_2)
	A ₂ and B	None (rarely α_1)

M, N, and P antigens (agglutinogens) Other agglutinogens, designated M, N, and P have been discovered more recently (1928) by Landsteiner and Levine. Approximately 50 per cent of white persons have both M and N (designated MN). Of the remainder about 28 per cent have M alone and 22 per cent have N alone. The plasma only very rarely contains antibodies (agglutinins) for these factors, but an anti-M or an anti-N agglutinin may develop in the blood of the recipient if he has been transfused repeatedly with blood from a donor belonging to the other subgroup. A reaction is not likely to occur, however, since the agglutinin is rarely, if ever, of sufficiently high titer to cause agglutination.⁴

⁴ Each of the M, N and MN groups has been subdivided by Sanger and Race by tests with a rare agglutinin discovered by Walsh and Montgomery. Serum containing this agglutinin agglutinates 30 per cent of MN blood, 19 per cent of M and 8 per cent of N group.

THE RH (RHESUS) FACTOR, AGGLUTINOGEN OR ANTIGEN

For a number of years fatal accidents had been reported following transfusions with blood which, in so far as the A-B-O and M, N and P groups were concerned, should have been compatible. In 1939 Levine and Stetson reported on a patient suffering a nearly fatal reaction from a first transfusion following her delivery of a dead fetus. They attributed the reaction to isoimmunization of the mother to an antigen derived from the fetus. The serum of the patient agglutinated the red cells of the great majority of white persons tested.

The first experimental work leading to the discovery of an antigen other than those already known was that of Landsteiner and Wiener in 1940. They found that when blood of the rhesus monkey (*Macaca mulatta*) was injected into a rabbit or guinea pig an antibody (agglutinin) was formed in the rodent's blood, which agglutinated not only rhesus red cells but the corpuscles of human blood, as well. An agglutinin similar to that produced by rhesus blood was discovered by Wiener and Peters in patients who had suffered severe transfusion reactions, though they had received blood which was apparently compatible. This antibody was shown to be independent of the A B O, M and N and P systems. Serum from these patients agglutinated the cells of over 80 per cent of white persons investigated. Wiener and Peters made a survey of similar transfusion accidents, and pointed out that in nearly all instances the patient had either received at least one transfusion some time before the one which had caused the reaction, or was a woman who was pregnant, or but recently delivered.

The similarity of the agglutinogen in human blood to that in rhesus blood, and the detection of the antigen by the use of a serum active against rhesus red-cells, led to its being named the rhesus or Rh factor by Wiener and Landsteiner.

The Rh factor is contained in the blood of 83 to 86 per cent of white persons; they are called Rh-positive (Rh+). The remaining 14 to 17 per cent are called Rh negative (Rh-). An antibody (anti-Rh agglutinin) is produced in the blood of an Rh-negative subject (as in the blood of a rabbit which has been injected with rhesus blood), if he should receive a transfusion of Rh positive blood. This process of isoimmunization takes about 12 days. Should he receive a second transfusion after this time a severe, perhaps a fatal reaction results,

the transfused cells undergoing agglutination and hemolysis. But, as indicated above, a pregnant woman may become immunized by the blood of an Rh+ fetus if she herself be Rh- (see Hemolytic Disease of the New-born, p. 49), and suffer a dangerous reaction if transfused with Rh positive blood.

This field has been greatly complicated by the later discovery of other Rh factors. There are now eight Rh antigens. Only six of these are at all common in the erythrocytes of Europeans, and over 95 per cent of all dangerous Rh reactions, and of cases of hemolytic disease of the new-born are due to the antibody of the one first discovered, it being the most powerfully antigenic. Following the notation of Fisher, this first-known Rh antigen is called D, and its antibody (produced by an Rh-negative person who has received a transfusion of Rh-positive blood) is known as anti-D. The other antigens are also designated by letters, and their antibodies by the letter preceded by the prefix anti-. Thus the six antigens are called C, c, D, d, E and e, the respective antibodies are known as anti-C, anti-c, anti-D, anti-d, anti-E and anti-e. If the red cells contain the antigens C, D and E, or any of them, the blood is Rh positive, but C or E is rarely present without D. Rh-negative blood is lacking in C, D and E antigens but contains the other three, c, d and e. These latter are but slightly antigenic, only c, except in rare instances, is capable of causing the production of an antibody in an Rh-positive person, and so of causing a reaction. See also *The inheritance of Rh factors* on page 48.⁶

In order to be classed as Rh negative, all Rh factors must be absent from the blood. Nearly 100 per cent of other races contain the Rh factor either alone or in combination, i.e. about 100 per cent are Rh positive.

Blocking antibodies, glutinins, univalent antibodies. Antibodies, which, through not causing demonstrable agglutination of Rh-positive cells suspended in saline, unite nevertheless with such cells and thus prevent their

⁶ To those who are not engaged in research in this field, or not in constant touch with laboratory procedures respecting it, the different terminologies have caused much confusion. The CDE-cde system of Fisher and Rice has been used here, since it is perhaps easier to remember. But Wiener's method of notation is used by many. In this system the six antigens are named, rh' (= C), Rh₀ (= D), rh" (= E), hr' (= c), Hr₀ (= d), hr" (= e). The reversal of letters (Rh to Hr) indicates, as do the capital and small letters of the CDE-cde system, the reciprocal relationship of the antigens so designated.

union with the agglutinating antibodies, are called "blocking" or "incomplete" antibodies, they are sometimes also qualified by the terms, "inhibiting", "non-agglutinating", "conglutinating", "late", or "mature" antibodies. If, however, the Rh-positive cells are suspended in serum, a solution of beef albumin (2 per cent cells in 20 per cent albumin), acacia or gelatin, agglutination occurs. A test for blocking antibodies consists, therefore, of placing a sample of the unknown serum in each of two tubes and adding a saline suspension of Rh-positive red cells to one and a suspension of cells suspended in beef albumin to the other. Clumping of the cells occurs in the second tube, but not in the first. Another test is to mix the unknown serum with a saline suspension of Rh+ cells and, if no agglutination occurs, to add a potent anti-Rh serum, the absence of agglutination indicates the presence of blocking antibodies in the unknown serum, which have prevented the action of the agglutinating antibodies in the second sample of serum.

Blocking antibodies appear later in life than the ordinary agglutinins, but they are more powerful than the latter and persist in the blood for many years once they have been formed.

In order to explain the action of blocking antibodies, Wiener and colleagues have offered the theory that of the two types the blocking antibody is of smaller molecular size than the agglutinating, and possesses only one side-chain or haptene, he therefore calls them univalent. The larger molecule of the agglutinating antibody is bivalent, possessing two haptenes, and can therefore unite with two red cells and cause clumping, whereas the molecule of the blocking antibody can unite with the antigen of but a single cell, but in doing so it blocks the action of the agglutinating antibody.

A suggestion which helps to explain the inability of the blocking antibody to cause agglutination of cells suspended in saline, but not of cells suspended in albumin, is that the saline alters the surface contours of the red cell, which can then become attached to only one haptene of a bivalent antibody—there being necessarily no difference in the molecules of the two types of antibody. On the other hand, there is some evidence that a blocking antibody is an agglutinating molecule altered in some way by the action of the saline.

At any rate it seems clear that in the body, since the cells are suspended in serum, the so-called blocking antibodies do cause agglutination. It is only when tests are made with saline as the suspending medium that they do not appear and thus mask an agglutinating reaction.

Detection of the Rh factor in blood

The Rh factor is detected by means of anti-Rh serum. The human anti-Rh serum most easily obtained is anti-D, which reacts with about 85 per cent of the bloods of white persons and is generally employed. This serum

contains anti-D and usually anti-C as well. It is an advantage if anti-c is present, otherwise blood containing c antigen alone would be thought to be Rh-negative. As mentioned elsewhere, e is rare. The only source of human anti-Rh serum is the blood of a woman who has recently given birth to an erythroblastotic child (p 49), and in many such women the agglutinin titer is too low to be satisfactory. The anti-A and anti-B agglutinins must of course first be neutralized. This is done by adding Witebsky's specific substance (p 45) AB to the serum, or by adding cells of A or B group, the anti-A or anti-B agglutinin being taken up by the added cells leaving the Rh antibody. Diamond and Denton's method of determining the Rh factor is briefly as follows.

The unknown cells, after being twice washed in saline, are mixed in a 20 per cent solution of beef albumin (to reveal the presence of blocking antibodies) to make a 2 per cent suspension of cells. Two drops of this suspension and 1 drop of the test serum are mixed in a small tube, the mixture is incubated at 37°C. Clumping of the cells, which is usually apparent in from 5 to 10 minutes, indicates the presence of the Rh factor. An alternative method using microscope slides as in A-B-O testing may be employed.

HEREDITARY TRANSMISSION OF THE BLOOD GROUPS

It is sometimes thought that close relatives must have compatible bloods, and that consequently a child could be transfused without harm with the mother's blood. But there is no gross mixing of fetal and maternal bloods, each retains its especial serological characteristics. It is true, on the other hand, that the iso-agglutinins are established much later after birth than the iso-agglutinogens, and may be in very low concentration in the new-born (whose blood would thus correspond to Group AB). It has therefore been suggested that the mother's blood no matter to what group it belonged would not undergo agglutination when transfused into the baby, but this is a decidedly dangerous assumption, for iso-agglutinins are not always in low concentration at birth.

The blood once established in its group remains unchanged throughout life. Even after a large number of transfusions, blood retains its original serological characteristics.

Inheritance of the main groups, A-B-O. According to Bernstein's triple allelomorph theory, the inheritance of the main blood groups is by three pairs of allelomorphous genes A, B and O, for which only one locus is provided on each chromosome of the chromosome pair. A and B represent the corresponding agglutinogens A

and B, O represents the absence of A and B. Since only one gene at a time can be situated on a locus, there are only 6 possible combinations or genotypes. Three genotypes are homozygous and three heterozygous. A and B are Mendelian dominants, O is recessive.

When O, which represents a recessive character, is transmitted both by the father and mother, then the child must have the genotype OO and be of group O. For A or B to be inherited at least one parent must possess and transmit an A or a B gene, respectively. If the child receives A from each parent (AA), or A from one and O from the other (AO), it will belong to group A. Similarly, if B is transmitted by each parent (BB), or B from one and O from the other (BO), it will belong to group B. The inheritance of A from one parent and B from the other (AB) will produce the blood group AB. It is evident then that a child cannot belong to group O if either parent belongs to group AB, for the child must receive at least one dominant gene A or B. Nor can a child be of any group but O (genotype OO) if both parents are of that group, for neither has an A or a B gene to transmit. Again, if only one parent belongs to group O, the child must belong to group A (genotype AO) or to group B (genotype BO) and cannot possibly be of group AB (see table 9).

The inheritance of Rh antigens. The theory of Fisher postulates three Rh genes C, D and E or their allelomorphs, c, d, e, closely associated or linked in the chromosomes. C, D and E are Mendelian dominants, c, d and e are recessive. The single Rh chromosome in the sex cell of the parent or in each chromosome of the pair in a body cell has three positions or loci, each of which can be occupied by a C or c gene, but not by both together, and similarly for D and d, or for E and e. The

gene combination is known as the genotype of the individual. Upon the genotypes of the parents will depend, of course, the combination of genes that the child will inherit. For example, if each chromosome of the pair of Rh chromosomes in the body cells of the father carries a D gene, then in respect to his Rh constitution (disregarding C and E) he may be designated DD (homozygous), and the Rh chromosomes of all his sex cells will contain a D gene, which he will transmit to his offspring. If the father's Rh constitution is Dd (heterozygous) half of his sex cells will have a D gene and half a d. The child may receive either. Similarly in the case of the mother, who, depending upon her Rh constitution, will transmit any of the antigenic genes. Therefore, it is mathematically possible for a child to have one of a number (36) of genetic combinations. All but five are very rare. These may be expressed as follows, CDe/cde, CDe/CDe, cde/cde, CDe/cDE, cDE/cde. The child would be Rh positive if he had any one of these genotypes except the third (cde/cde) which would make him Rh negative.

Race and his associates found that in a survey of the bloods of some 2000 English persons the genotype CDe/cde was the most common (32.7 per cent). CDe/CDe came next in frequency (17.7 per cent) and the other three in descending order as given (15.1, 11.9 and 11.0 per cent). These five genotypes, therefore, account for nearly 90 per cent of the total.

The CDE-cde notation is rather cumbersome, so an abbreviated, shorthand terminology, based upon that mentioned in foot note, p. 46, is usually employed, thus with CDE-cde equivalents is shown below:

$$\begin{aligned} \text{CDe/cde} &= R_1r \\ \text{CDe/CDe} &= R_1R_1 \\ \text{cde/cde} &= rr \\ \text{CDe/cDE} &= R_1R_0 \\ \text{cDE/cde} &= Rr \end{aligned}$$

Fisher and Race formulated their theory when only one Hr (H⁰) antigen was known, the other two, which were discovered shortly afterwards, were predicted by the theory.

The genetic relationships of the Rh Hr factors as well as of the A-B-O and M-N groups are of great importance in certain phases of forensic medicine, e.g., disputed paternity, and the identification of blood stains in cases of murder. Knowledge of the A-B-O, Rh and other serological characteristics of different peoples is also of value in ethnological studies. The greater the number of serological types into which the population can be divided the finer will be the discrimination which can be made between two blood specimens. Thus, should both parents be Rh-negative (genotype H⁰H⁰, cde), the child cannot receive an Rh factor, so must also be Rh-negative, or if a parent is homozygous for any Rh factor, at least one such factor must be transmitted to the child who cannot therefore be homozygous for the reciprocal H factor.

TABLE 9

Summary of published studies on the heredity of the blood groups

(From Wiener, after v. Dungern and Hirsfeld)

PARENTAL COMBINATIONS	NUMBER OF FAMILIES	NUMBER OF CHILDREN IN EACH BLOOD GROUP			
		O	A	B	AB
O × O	1563	3772	(14)	(9)	0
O × A	2903	2707	3749	(10)	(1)
A × A	1385	556	2538	0	(2)
O × B	1456	1418	(7)	1831	(1)
B × B	554	203	(1)	1009	0
A × B	1400	605	957	771	848
O × AB	530	(8)	633	646	(3)
A × AB	455	0	533	247	312
B × AB	323	(2)	183	406	232
AB × AB	59	0	28	36	65

Exceptions to the rule are shown in parentheses, probably illegitimates.

The inheritance of the M and N factors is of a similar character (two sets of closely linked allelomorphs) to that of Rh-Hr. There are but three possible combinations, genotypes MM, MN and NN, corresponding, respectively, to the phenotypes M, MN and N. It follows that a parent with an M type (MM) cannot have an N type child and, vice versa, an N type parent cannot have a child with blood of the M type.

HEMOLYTIC DISEASE OF THE NEW-BORN (ERYTHROBLASTOSIS FOETALIS)

Hemolytic disease of the new-born or Erythroblastosis foetalis is a blood disorder which, if severe and untreated, ends fatally, death of the fetus may occur in the uterus. It is usually due, as first suggested by Levine and his associates, to isoimmunization of an Rh-negative mother by a Rh-positive fetus, the fetal blood characteristic having been inherited from the father. Surveys of the serological characteristics in this disease have revealed that in approximately 93 per cent of cases the mother was Rh-negative and the infant Rh-positive.⁶ The Rh-antibody produced in the maternal blood crosses the placenta and destroys the fetal cells. The disease is characterized by jaundice (*Icterus gravis neonatorum*), a progressive hemolytic anemia with large numbers of primitive erythroblasts in the fetal blood, enlargement of the liver, and frequently massive edema of the entire body with ascites (*hydrops foetalis*). In some instances degenerative changes with the deposition of bile pigment are found in the basal nuclei of the brain (*kernicterus*) which are believed to be due to thrombi in the minute vessels in the neighborhood.

Of the 7 per cent of cases of hemolytic disease of the new-born which cannot be accounted for by incompatibility with regard to the Rh factor, some are believed to be due to incompatibility between mother and fetus with regard to the main groups A or B. Incompatibility with respect to the A-B-O system may also be responsible for some cases of habitual abortion.

It is evident that transfusion of a mother either before or after she has given birth to an Rh-positive infant will be attended by the gravest danger unless precautions are taken to ensure that the donor's blood is compatible, not only with regard to the A-B-O groups, but also with respect to the Rh factors. After isoimmunization, a high titer

of the Rh-antibody persists in the body for a considerable time, perhaps for life.

All Rh-positive children born of Rh-negative mothers are not erythroblastotic. The incidence of the disease in the white population is about 0.5 per cent of births, whereas the matings of Rh-positive men with Rh-negative women is 13 per cent. There are several reasons for the relatively low incidence of the disease. The first of these is the tendency toward small families combined with the fact that the disease is rare in a first child, unless an Rh-negative woman has been artificially immunized by having had a transfusion of Rh-positive blood for any reason previously, whether pregnant or not. Isoimmunization becomes more effective with each succeeding pregnancy. Secondly, a large proportion of Rh-mothers (95 per cent) seem incapable of developing a high titer of Rh-antibody. Finally, the Rh-positive father, if heterozygous for the Rh antigen, will not have children who are all Rh-positive, half will be Rh-negative. The incidence of the disease in certain races, e.g. Japanese, Chinese, Negroes⁷ and American Indians is very rare, since they are nearly all (99–100%) of one group with respect to the Rh factor, namely, Rh-positive.

The manner in which isoimmunization is effected—Some difficulty has been encountered in attempting to explain how isoimmunization is brought about for the reason that, like other agglutinogens, the Rh factor is attached to the red cell, and it has been generally believed that the placenta offered a perfect barrier to prevent any mixing of fetal and maternal bloods. Yet to explain isoimmunization of the mother, it must be assumed that red cells or fragments of red cells pass from fetus to the maternal circulation. It appears that isoimmunization commences in the second half of pregnancy when the chorionic villi become intimately related to the maternal blood sinuses, a syncytium consisting of a single layer of cells alone separating the two circulatory systems. It is easily conceived how small breaks in this system would be sufficient to permit small fragments of Rh-positive cells to pass to the mother's blood. The extremely small quantities of an antigen which are required for antibody production is well known, and minute amounts of Rh antigen entering the maternal circulation over a period of weeks would be quite

⁶ The univalent (blocking) antibodies rather than the bivalent are mainly responsible since they cross the placental membrane much more readily.

⁷ The incidence in American Negroes is approximately 91 per cent, but this relatively low figure is probably due to an admixture of blood of the white race.

effective in inducing isoimmunization.* Since it is in solution, there is no difficulty in accounting for the transference of the anti-Rh agglutinin formed in the mother's blood across the placenta.

The treatment of hemolytic disease of the newborn consists in bleeding the infant and at the same time replacing its blood with Rh-negative blood (*exchange or exsanguination transfusion*). If Rh-positive blood is employed, the red cells are rapidly destroyed by the antibodies in the infant's blood or attached to the tissue cells. It is now very widely agreed that Rh-negative blood is preferable for transfusion, though it is still argued by some that by the use of Rh-positive blood and thereby fixing the destructive antibodies these are more quickly eliminated from the body.

Attention should again be called to the fact that anti-Rh agglutinins are not ordinarily present, in any human serum, that is, without a previous active isoimmunization. That being so, danger of a reaction can only arise in an Rh negative man who has previously received a transfusion of Rh positive blood, or in an Rh-negative woman who has had a transfusion of such blood or has given birth to an Rh positive child. It is also said that only from 2 to 4 per cent of Rh-negative persons become immunized after receiving a transfusion of Rh-positive blood. This accounts for the fact that there were not more transfusion fatalities in the years before the discovery of the Rh factor.

Cold hemo-agglutination This is the agglutination which occurs outside the body between red cells and their own serum, or any other human serum, when the temperature is reduced much below 25°C. It is due to a non specific antibody called a "cold agglutinin." This agglutinin is present in low titer in 95 per cent of normal persons. The phenomenon, sometimes also called autoagglutination, differs from exaggerated rouleaux formation (pseudoagglutination) in that cold is the essential precipitating factor. Cold agglutination is a feature in a number of diseases, the agglutinin being present in high titer. Such diseases are bronchopneumonia, atypical pneumonia, influenza and other acute respiratory infections, Raynaud's disease, hemolytic anemias, paroxysmal hemoglobinuria, acute bacterial infections, cirrhosis of the liver and a number of other pathological states. Occurring *in vivo* it may lead

to gangrene of the tips of the fingers and toes. The agglutination is reversible, the cells becoming separated again when the suspension of red cells and serum is warmed. The antibodies responsible for cold agglutination are unrelated to the A-B-O, M-N or Rh systems or to rouleaux formation.

The importance of cold agglutination lies in the confusion which may arise in testing for incompatibility between two blood samples—isoagglutination.

(2) HUMAN SERUM AND PLASMA

Many of the drawbacks inherent in the use of whole blood for transfusion are obviated by substituting human serum or plasma. These blood derivatives when pooled from a number of donors can be transfused indiscriminately without regard for the blood groups. Furthermore, liquid plasma or serum if collected under strictly sterile conditions can be stored under refrigeration for an indefinite period. Plasma kept in the frozen state is highly recommended by Strumia and McGraw as being safer than storage in liquid form. Plasma or serum dried *in vacuo* from the frozen state by the *desicc* process of Flosdorf, Stokes and Mudd is the preparation *par excellence*, especially for use in the armed services. It can be stored without refrigeration, is easily transported and the dangers of bacterial contamination or of deterioration are at a minimum. But like whole blood it may transmit the virus of homologous serum jaundice. When required for use, distilled water, equal in volume to that which had been removed in the drying process, is added to the desiccated material.

(3) SOLUTIONS OF COLLOIDS

The limited availability of whole human blood or of its derivatives, plasma or serum, has inspired a search for a blood substitute which would resemble as closely as possible the physical properties of plasma. The requirements of an artificial transfusion material are several: (a) The molecules (or particles) of the substance must be of such a size that its solution will not leave the circulation too freely. (b) The solution must have an osmotic pressure and a viscosity approaching as closely as possible these properties of whole blood, such qualifications depend upon molecular size and shape. (c) It should be isotonic with the contents of the corpuscles. (d) It must, of course, be non-toxic and innocuous in every respect. (e) It should not hinder normal plasma protein production. In

* Levine has calculated from experiments on rabbits that no more than 0.067 cc. of packed Rh positive cells would be required to pass from the fetus to the maternal circulation to cause isoimmunization.

addition, the material should be readily available in large quantities, preferably cheap, capable of being sterilized by simple means, and of being quickly prepared for use. Provided a material is suitable for transfusion in the foregoing respects there appears to be no valid objection to the use of some fluid other than blood or its derivatives to fill the vessels after hemorrhage. The properties which render a fluid suitable for transfusion are physical rather than chemical. Serum or plasma apparently possesses no advantage by virtue of any biochemical characteristics which it possesses. Whole blood is superior to either of these in severe traumatic or hemorrhagic shock.

Gum acacia in a 6 per cent solution in saline was introduced by Sir William Bayliss as a transfusion fluid during the first World War and proved highly successful. The osmotic pressure and viscosity of this solution closely resemble those of plasma. The molecules of gum acacia aggregate into particles comparable in size to those of the plasma proteins and do not escape freely through the capillary membrane.

However, experience since the first World War has revealed some undesirable features of gum acacia. Gum acacia is a polysaccharide which the body cannot metabolize. It, like other foreign materials which cannot be disposed of in other ways, is taken up by the reticulo-endothelial elements of the liver and elsewhere. In animals, swelling and vacuolization of the hepatic cells have been observed following the transfusion of this material. The liver may become greatly enlarged after repeated transfusions with gum acacia and sometimes shows areas of necrosis. The hepatic function of producing plasma protein is seriously depressed, the concentration of protein in the plasma is lowered following large transfusions of gum acacia and may remain below normal for a considerable time after this blood substitute has been administered.

Isinglass (collagen prepared from the swim-bladders of certain species of fish, e.g., sturgeon, hake, etc.) has been advocated by Taylor and Waters, in a concentration of 6 per cent in physiological saline, as a blood substitute. Isinglass as now prepared is free from anaphylactic or pyrogenic action. It has been used clinically and found to be effective and quite safe. It has been shown both in patients and in animals to raise the blood pressure to normal after severe hemorrhage and to maintain it at this level for several hours. After

its injection the blood shows progressive dilution indicating that not only is it retained within the circulation but exerts an osmotic effect which "draws" interstitial fluid into the vessels. It does not appear to have any depressing effect upon plasma protein production by the liver and has a technical advantage over animal gelatin in being perfectly fluid at room temperature. The molecular weight of isinglass, after autoclaving, is between 18,000 and 30,000.

Animal gelatin, which is closely similar to isinglass chemically, or *dextran*, a polysaccharide with a large molecule, have been used clinically with satisfactory results.

Among other colloid transfusion materials is *hemoglobin-Ringer* solution which is under investigation mainly by Amberson and his colleagues.

The *albumin* fraction of human serum as developed by Cohn, Janeway and their associates is, except for its cost, an ideal transfusion material when only a cell-free fluid is required. It may be given in concentrated solution or in the same concentration as in normal plasma. *Bovine plasma*, if it could be rendered quite safe yet still retain its other qualifications for a transfusion material, would solve the problem of a physiological and relatively cheap blood substitute for those emergencies in which it is not necessary to furnish red-cells. The blood proteins of another species are, of course, intensely anaphylactogenic, and several attempts have been made to remove this property (*despecciation*) from bovine plasma or albumin. Masson has described a method involving the use of formaldehyde and heat to despecciate calf plasma. But such procedures which denature the proteins tend at the same time to reduce their osmotic pressure to a degree which greatly lessens their value for transfusion.

C 446 (4) SOLUTIONS OF CRYSTALLOIDS

The small molecules of salt or of glucose pass freely through the capillary wall. They exert a negligible osmotic effect, the injected fluid is therefore not retained in the circulation. For this reason such fluids though capable of raising the blood pressure temporarily are quite unable to maintain it for any considerable length of time. Indeed they may do serious harm, for the transfused fluid as it leaks into the tissues carries plasma protein with it. This is particularly likely to occur in the case of saline. Transudation of fluid into the tissue of the lung—pulmonary edema—may result. When, on

the other hand, dehydration of the tissues and loss of blood *water*, as shown by the concentration of the plasma proteins, are prominent features, then the subcutaneous or intravenous injection of saline (with the addition of glucose) or water by mouth, would appear, from physiological principles, to be a logical procedure. In the dehydration resulting

from chloride depletion (p. 24) sodium chloride is clearly indicated. A solution of this salt serves not only to supply fluid but to replenish the base the loss of which is such an important factor in the development of the dehydrated state. It is also of great value in correcting the hemoconcentration which results from extensive burns.

CHAPTER 6

HEMOGLOBIN

See also Regeneration of Blood, ch 8

Hemoglobin is the coloring matter of the erythrocytes, and the chief function of the red cell is to store this pigment and carry it around the circulation. About 10 grams of hemoglobin pass through the lungs per second. It takes up a comparatively large load of oxygen which it carries to the tissues. One hundred cubic centimeters of water at the temperature of the body and exposed to an oxygen pressure of 100 mm Hg absorbs a third of a cubic centimeter of the gas. One hundred cubic centimeters of blood, on the other hand, at the same temperature and pressure will take up about 20.0 cc—that is, 60 times more. The difference is due to the hemoglobin. The total amount of blood in the human body will hold approximately 1200 cc of oxygen (200 cc per liter of blood). This quantity of oxygen is used by the tissues in 5 minutes or so during rest and in a fraction of a minute during muscular exertion. In the absence of hemoglobin, the entire duty for the carriage of oxygen would have to be performed by the plasma, and in order that this should be able to absorb the necessary amount of gas, it would have to be increased at least 60 times in amount. As pointed out by Barcroft, the circulating fluid, instead of being about 6 liters or $\frac{1}{11}$ of the body weight, would then need to be over 350 liters. That is, more than five times the bulk of the solid tissues.

Besides serving as a carrier of oxygen, hemoglobin plays an important rôle in regulating the acid-base balance of the blood (ch 13), and in the carriage of carbon dioxide (ch 32).

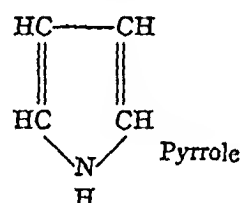
THE CHEMICAL CONSTITUTION OF HEMOGLOBIN THE "STONES" FROM WHICH ITS MOLECULE IS BUILT

Hemoglobin is a conjugated protein consisting of an iron-containing pigment portion combined with a protein of the histone class called *globin*. The hemoglobin complex, when its globin is in the natural state, forms a loose combination with oxygen (oxygenation)—the iron being in the ferrous state (Fe^{++}). Under certain abnormal conditions it forms a stable compound with oxygen (oxidation, i.e., the production of a true oxide)—

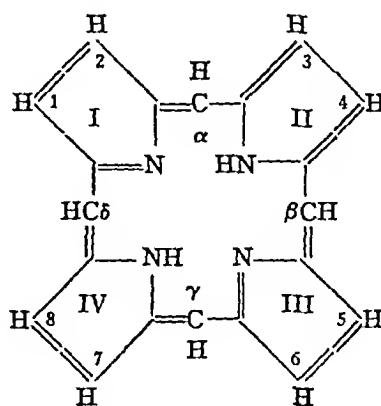
the iron being in the ferric state (Fe^{+++}). A more detailed consideration of hemoglobin structure follows.

PORPHYRINS are pigments which, either alone or as the basis of more complex compounds, are found throughout plant and animal life from the highest to the lowest forms. A porphyrin is the pigment basis of chlorophyll—the green coloring matter of plants. One is found as a brown pigment in the shells of many eggs, and also in the dark line running down the back of the earthworm. On the other hand, when conjugated with other substances porphyrins are the basis of the blood and tissue pigments of various animals.

The basic nucleus or framework upon which all *porphyrins* are built consists of 4 pyrrole rings



linked together in a larger ring-like structure by 4 methene couplings. This parent structure is called *porphin* and is shown below —



Porphin, $\text{C}_{20}\text{H}_{14}\text{N}_4$

By substitution of its eight hydrogen atoms by various groups—ethyl, methyl, vinyl or propionyl—the several porphyrins in nature are produced. In the animal body it is believed that synthesis occurs in two stages. First, substances composed of

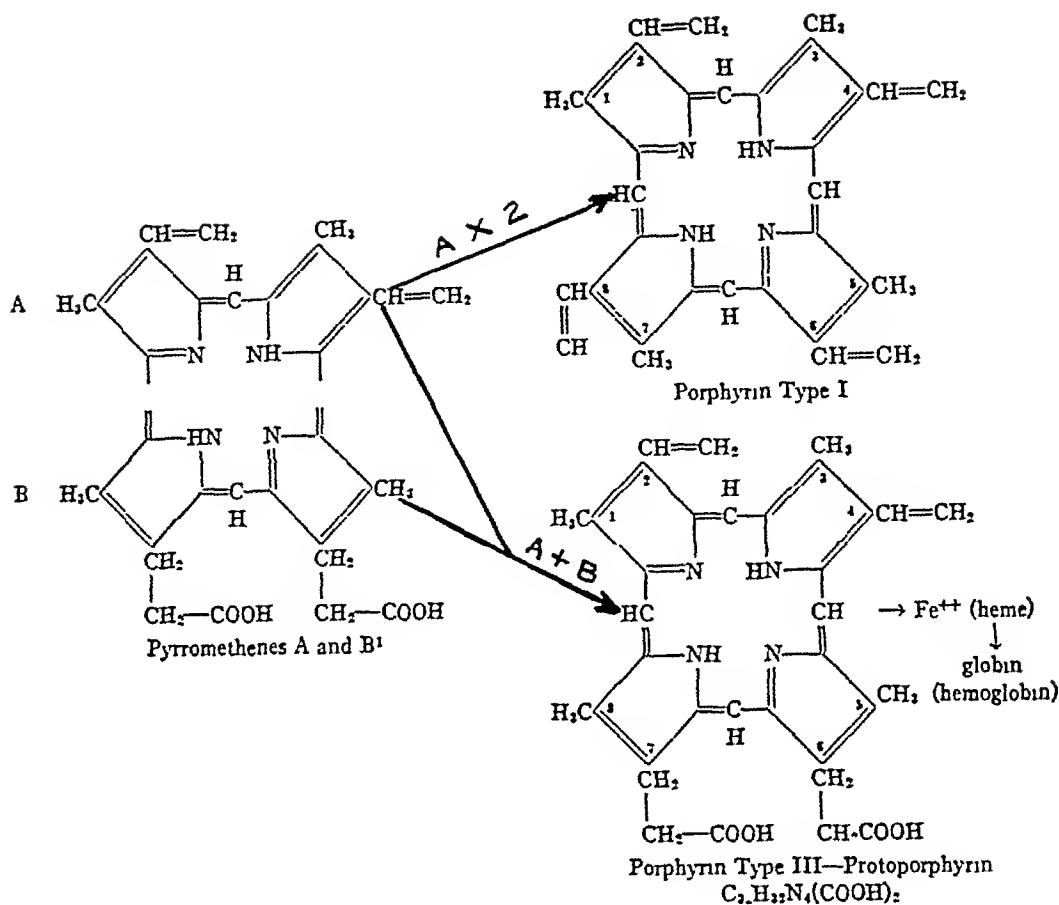


FIG 61

two pyrrol groups only are formed and known as *pyromethenes*. These are of two kinds. In one form, A in figure 61, the four hydrogen atoms are substituted by two methyl (CH_3) and two vinyl ($-\text{CH}=\text{CH}_2$) groups. In the other form, B, the hydrogen atoms are substituted by two methyl and two propionyl ($-\text{CH}_2-\text{CH}_2-\text{COOH}$) groups. When two pyromethenes of the first type unite, a type I porphyrin is produced. When an A and a B form unite, type III porphyrin is formed. Much smaller quantities of the type I pigment are produced in the body than of type III, but in health a constant ratio exists between the two. Type III porphyrin is the pigment of hemoglobin, myoglobin and certain respiratory enzymes. It is called *protoporphyrin* and is isomeric with *oöporphyrin*, the pale brown pigment in the eggshell of the domestic hen (see fig 61).

¹ A type II porphyrin might possibly be formed by the union of two B types of pyromethenes, but such has not been found in nature.

Protoporphyrins may be looked upon as pigment complexes in which the eight hydrogen atoms of the porphyrin nucleus are replaced by three of the groups mentioned above, namely by 2 vinyl, 4 methyl, and 2 propionyl. There are some 15 different isomeric protoporphyrins, the one present in hemoglobin being designated protoporphyrin No. 9. Its vinyl groups are at positions 2 and 4, its methyl groups at 1, 3, 5 and 8, and its propionyl groups at 6 and 7.

Small amounts of free protoporphyrin are present in the erythrocytes and larger amounts in reticulocytes. The presence of this pigment is probably responsible for the red fluorescence exhibited by the latter when viewed under ultra-violet light.

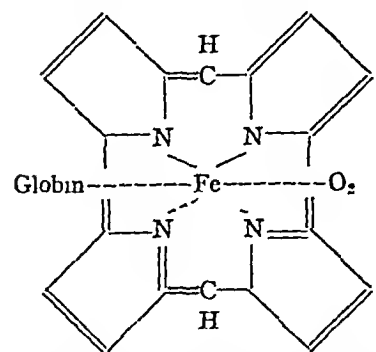
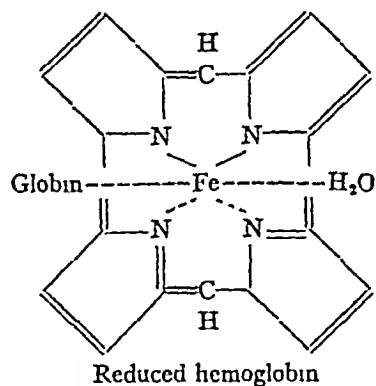
The two animal porphyrins (I and III) remain distinct chemical entities under all circumstances, being synthesized separately and never undergoing interconversion. The conception of a "dualism of the porphyrins" was first advanced by Hans Fischer and has been supported in more modern

times by Dobriner, Watson and other workers in this field

In health, Type I porphyrin and sometimes Type III are excreted in minute amounts in the urine after being converted to *coproporphyrin* (p 60) Coproporphyrin I, but not Type III, is excreted through the bile *Uroporphyrin* is not found in normal urine but is excreted in the urine in congenital porphyria

Metallo-porphyrins Porphyrins are capable of forming compounds with various metals A pigment found in the feathers of a certain South African bird (turaco), for example, and known as *turacin*, is a porphyrin combined with copper. Other metalloporphyrins have been prepared, namely, those of cobalt, nickel, silver, manganese, tin, zinc, etc.² Protoporphyrin combined with iron forms the metallo-porphyrin of the blood pigment. For instance, if an Fe atom be attached to protoporphyrin we get the iron-porphyrin compound of hemoglobin This is called *heme* ($C_{34}H_{32}N_4O_4Fe$) As we shall presently see, however, heme is not peculiar to hemoglobin but is a constituent of other respiratory substances

The atom of iron, in the ferrous state, is believed to be attached to the porphyrin group in the manner represented below



² In chlorophyll the porphyrin is combined with magnesium

In the hemoglobin molecule four heme molecules are attached to the globin molecule

Heme is capable of combining with various proteins or nitrogenous substances, e.g., albumin, ammonia, pyridine, nicotine, etc. Such compounds are called *hemochromogens*. When globin is the protein with which heme is combined the resulting hemochromogen is that forming the basis of the blood pigment of vertebrate life. Hemoglobin is, therefore, an iron + porphyrin + globin compound

THE HEMOCHROMOGEN OF HEMOGLOBIN When hemoglobin is treated with alkali, its characters are changed. Its spectrum differs from that of hemoglobin. It still is capable of combining with oxygen, but unlike hemoglobin it does not readily release the gas again. For many years this modified hemoglobin, to which the term *hemochromogen* was first applied, was thought to contain only the pigment fraction (porphyrin + iron) of the hemoglobin molecule, the globin supposedly had been "sheared" off. The work of Anson and Mirsky, however, has shown that this assumption is incorrect. No separation of globin from the pigment fraction occurs. Hemochromogen is a porphyrin + iron + globin compound, just as is hemoglobin. The globin, however, has been denatured by the action of the reagents and to this denaturation the properties of hemochromogen whereby it differs from hemoglobin are ascribed. The iron is in the ferrous state (Fe^{++}). Hemochromogen consists therefore of denatured globin combined with the pigment complex $C_{34}H_{32}N_4O_4Fe$ (heme).

When hemochromogen is exposed to oxygen a true oxide is formed. This is called *cathemoglobin*. The heme is oxidized—*oxidized heme*—the iron being in the ferric state (Fe^{+++}). Cathemoglobin is therefore oxidized heme ($C_{34}H_{32}N_4O_4FeOH$) united to de-natured globin.³ *Methemoglobin* is oxidized heme combined with native globin.

HEMIN ($C_{34}H_{32}N_4O_4FeCl$) is the hydrochloride of heme and is prepared by heating oxyhemoglobin with glacial acetic acid and a minimal amount of sodium chloride. Upon cooling, reddish brown prismatic crystals of hemin separate out. The detection of these, which are frequently referred to as Teichmann's crystals after their discoverer, is used as a test for blood in suspected stains. When hemin is treated with caustic soda oxidized heme is obtained. When the latter is treated with a

³ Oxidized heme is called *hematin* by some and heme is referred to as *reduced heme* by Anson and Mirsky

weak acid the iron is split off and protoporphyrin remains. Hemin or blood itself when treated with concentrated mineral acids in the presence of oxygen yields hematoporphyrin (p. 60).

THE DISTRIBUTION OF HEME IN NATURE

Heme is almost universally distributed throughout the animal and vegetable kingdoms. Respiratory pigments with this porphyrin iron compound as their common basis are found in the lowest forms of plant life as well as in the highest species of animals.

Cytochrome is a heme compound which is widely distributed in the tissues of plants and animals. It is present in certain aerobic bacteria, in yeast cells, in the onion, in worms, molluscs, crustacea, in the muscles of the bee's wing and in many other insects and their larvae. It is present in the muscles and other tissues of the large number of vertebrate species which have been examined. Cytochrome plays an important part in the oxidation system in the tissues (p. 380). It undergoes alternate oxidation and reduction but unlike hemoglobin is not autoxidizable or only slightly so (due to component *b*, see below). In order to take up oxygen it requires the aid of tissue oxidase (indophenol oxidase), in order to undergo reduction it requires the presence of dehydrogenases. These activate the hydrogen of organic molecules in the tissue cells which become hydrogen donors. The cytochrome acts as a hydrogen acceptor. In this way cytochrome, it is suggested, serves as an intermediary in the transference of oxygen, liberated from hemoglobin, to the oxidizable materials in the tissue cells. It may also through component *b* serve for the direct transference of oxygen. In the presence of cyanides, carbon monoxide (in the dark) or sulphides which poison the oxidase, the oxidation of cytochrome is inhibited. On the other hand, anesthetics which depress the action of dehydrogenases prevent its reduction. In either case the link in the chain of oxygen usage by the cells is broken.

Cytochrome is a mixture of three hemochromogens, they are referred to by Kellin as *a*, *b* and *c*. Of these only *b* is autoxidizable. The heme components of the hemochromogens are not all the same, there being two varieties. One of these is identical with that in hemoglobin, the other resembles that in chlorocruorin, see below. The nitrogenous compounds with which the hemes are combined are unknown. Cytochrome is identified in living tissues by its characteristic ab-

sorption spectrum. The cytochrome of bees' wing muscle shows four absorption bands at 6046, 5665, 5502 and 5210, Angstrom units (see p. 59), respectively. Its oxidation and reduction can be followed in the living cell by means of the microspectroscope, the bands becoming distinct when reduction occurs but almost disappearing when the substance is oxidized.

Heme in the free state, that is, uncombined with a nitrogen compound, has been discovered in many substances such as wheat flour and oatmeal where its presence had never been suspected. The fact that heme is in one way or another of such universal occurrence has prompted Barcroft to remark, "mankind has for countless centuries been eating, all unknowingly, the outstanding constituent of his blood." It has been frequently suggested in the past that chlorophyll which also is constituted of pyrrol rings, was the primitive pigment and that animals probably derived the pyrrol grouping for the manufacture of the pigment of their bloods from this green coloring matter in their diet. It is now seen, however, that heme is a much more ancient pigment since it is found in the most elemental forms of plant life in which chlorophyll does not exist. It is pointed out, however, on page 74, that there is little evidence that either of these pigments in the diet serves as a basis for hemoglobin synthesis.

Hemoglobin itself is by no means so widely distributed as is heme and the heme compound cytochrome, for it is confined to the animal kingdom. It is found in the blood of all vertebrates and of several invertebrates, e.g., worms, a certain snail, in the larvae of some but not in the body fluids of any adult insect. The blood pigments of different vertebrate species vary in their properties, the hemoglobin of the frog, for example, possesses a spectrum and oxygen dissociation curve different from that of mammals. The variability is due to minor differences in the globins to which the heme is joined and not to different hemes, which are the same throughout the vertebrate phylum.

Though many different porphyrins exist, only one differing from protoporphyrin has been discovered in nature as forming part of a hemoglobin-like substance, i.e., one in which the nitrogenous fraction is probably a globin. A hemoglobin-like pigment, greenish in color, is found in certain worms and is called *chlorocruorin*. It contains this other porphyrin of unknown structure combined with iron. *Helicorubin* is a respiratory pigment found in the gut and liver of the snail. It contains the same heme as hemoglobin as shown by the fact that if its non-pigmented fraction is re-

placed by pyridine the pyridine-hemochromogen so formed shows a spectrum identical with that of the pyridine-hemochromogen derived from hemoglobin. Chlorocruorin, however, since its heme is different, when treated similarly shows a different spectrum. The nitrogenous part of heliocorubin is unknown but presumably it is not globin.

Hemocyanin is a respiratory pigment which in certain crustacea and molluscs (king crab, octopus and snail) takes the place of hemoglobin. It is dissolved in the circulating fluid and not confined within cells. This substance contains copper instead of iron but the metal is not combined with a porphyrin as was believed at one time. This pigment is blue when oxidized and colorless when reduced.

SUMMARY OF HEMOGLOBIN AND MYOGLOBIN SYNTHESIS, AND OF CERTAIN COMPOUNDS OF HEMOGLOBIN AND HEME

Pyrrole $\times 2$ = *pyrromethene* (Type I or III)

Pyrromethene $\times 2$ = *porphyrin*, types I and III
(*protoporphyrin*)

Protoporphyrin + divalent iron (Fe^{++}) = *heme*

Heme (iron in ferrous state [Fe^{++}]) + various nitrogenous substances = *hemochromogens* of various respiratory pigments, e.g., *cytochrome*, *heliocorubin*, etc.

Heme (iron in ferrous state [Fe^{++}]) + *globin* = *hemoglobin* (see also scheme on p. 61)

Heme (iron in ferrous state [Fe^{++}]) + *denatured globin* = *hemochromogen of hemoglobin*

Oxidized heme (iron in ferric state [Fe^{+++}]) + *denatured globin* = *cathemoglobin*

Oxidized heme (iron in ferric state [Fe^{+++}]) + *globin* = *methemoglobin* (p. 58)

THE MOLECULAR WEIGHT OF HEMOGLOBIN

The pigment proper (heme) constitutes about 4 per cent, and the globin about 96 per cent of the hemoglobin molecule. The small but relatively heavy porphyrin-iron portion is floated, as it were, by the large protein fraction. If hemoglobin contained only 1 atom of iron which has a molecular weight of 56, then since the percentage of the metal in hemoglobin is 0.334,⁴ as determined by direct analyses, the minimum molecular weight of hemoglobin would be $(56/0.334) \times 100 = 16,800$.

⁴ This is Hufner's figure, but other investigators have obtained lower, others higher values, namely from 0.305 to 0.338 per cent. The average of values obtained by Bernhart and Skeggs in a number of analyses of hemoglobin crystallized from the pooled blood of twenty human subjects was 0.340. Using this figure in the calculations, a minimal molecular weight of 16,400 and an oxygen capacity of 1.36 ml per gram of Hb are obtained.

approximately Hufner obtained in fact a figure of 16,800 from osmotic pressure measurements.

Adair, however, from a study of the osmotic pressure of hemoglobin finds the molecular weight to be approximately four times greater, namely, 68,000. The molecule contains 4 atoms of iron. Using an entirely different but most ingenious physical method, namely, ultracentrifugation, Svedberg obtained a value almost identical with Adair's. Solutions of hemoglobin were centrifuged at great speeds for several hours so as to force the sedimentation of the hemoglobin molecule, which under ordinary circumstances of course does not separate from solution. When a solution is subjected to a centrifugal force 200,000 times that of gravity, a conflict between two opposing forces occurs. Centrifugal force tends to throw the particles down out of the solution, diffusion, on the other hand, tends to prevent their sedimentation. At certain determinable values of these two forces equilibrium will become established and the molecular weight calculated.⁵

Northrop and Anson also obtained a figure of around 68,000 from calculations based upon the rate of diffusion of hemoglobin through discs of alundum. Hemochromogen is probably a depolymerized form of hemoglobin with a molecular weight of 16,700, or double this value.

The hemoglobin is thought to be disposed near the surface of the red-cell, the molecules being cylindrical in shape, closely packed but not so intimately as to prevent free rotatory movement.

Human blood contains about 15 grams (14 to 16 grams) of hemoglobin per 100 cc. Since the proportion of iron in hemoglobin is 0.334 per cent, the quantity of the metal in 100 cc of blood is about 50 mg and in the total blood of the human body about 4-5 grams. The blood contains a small proportion of iron in addition to that combined with hemoglobin (p. 76).

THE ESTIMATION OF HEMOGLOBIN IN BLOOD
Several methods are available for determining the hemoglobin concentration of blood. But whatever the method used, the fundamental information sought is the oxygen capacity of the blood. This may be found directly by means of the Van Slyke apparatus, or the apparatus of Barcroft or of Warburg. Or, since the

⁵ Since the sedimentation equilibrium depends upon both sedimentation and diffusion, the molecular weight can be obtained from these determinations and calculation from the following formula given by Svedberg:

$$M = \frac{2 RT \log (C_2/C_1)}{(1 - V_p) \omega^2 (X_2^2 - X_1^2)}$$

Where M is the molecular weight, R the gas constant, T the absolute temperature, V the partial specific volume of solute, ρ the density of solution, C_1 and C_2 the concentrations at the distances X_1 and X_2 , and ω is the angular velocity.

hemoglobin molecule contains 0.334 per cent of iron, the hemoglobin in grams per 100 cc. can be calculated from an analysis of the blood for its iron content.⁴ Thus, if the blood contains 50 mg. of iron per 100 cc., its hemoglobin concentration is $(50/0.334) \times 100 = 15$ grams (approx.) per 100 cc. and the oxygen capacity $(15 \times 1.34 =) 20$ cc. (approx.) per 100 cc.

The most convenient method for chemical use is one based upon matching a sample of diluted blood with one of a series of permanent color standards. The best known methods of this type are the *Haldane-Gowers*, in which the hemoglobin is first converted to carboxy-hemoglobin, the *Gowers*, in which the blood sample after dilution and full oxygenation is matched colorimetrically with a picocarmine standard, and the *Sahli* method, which involves the conversion of the hemoglobin in the diluted sample to acid hematin.

The Sahli method as modified by Haskins will be described. Dilute HCl (0.2 N) is placed in a small graduated glass tube up to the mark 10.002 cc. of the blood under examination is delivered into the acid with a Sahli pipette which is rinsed by drawing the acid blood mixture into it twice and blowing out each time. The tube is kept at a temperature of about 60°C. for 7 minutes. When the full reddish brown color of the acid hematin (hemin) which is formed has fully developed, the tube is placed in a comparator alongside a standard color solution and carefully diluted with water until the two colors exactly match. The hemoglobin percentage is given by the figure on the tube at the level of the mixture. The temperature of the standard should be kept at around 20°C. while the color comparison is being made. Blood containing 13 grams of hemoglobin per 100 cc. gives a reading of 100 per cent. The blood of the average healthy person would therefore give a reading of 110 per cent. In the Haldane-Gowers method the color scale of the apparatus is also set too low, 100 per cent on the scale representing a hemoglobin concentration of only 13.8 grams per 100 cc. instead of the mean normal of approximately 15 grams.

Phillips Van Slyke and their associates have devised a method whereby the hemoglobin concentration can be determined from the specific gravities of blood and plasma. The calculation is made from the following equation—

$$\text{Hb in grams per 100 cc. of blood} = 33.9 \times \frac{G_b - G_p}{1.0970 - G_p}$$

33.9 is the presumed mean figure for grams of hemoglobin in 100 cc. of red cells, G_b is the observed specific gravity of the blood sample, G_p the observed specific gravity of the plasma, 1.0970 is the presumed mean value for specific gravity of centrifuged red cells of normal blood.

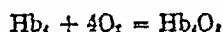
⁴ The estimation of hemoglobin by iron analysis is claimed by King and his associates to be the most reliable method, though even this method is not without error, for a certain small proportion of the iron in blood is not in combination with functioning hemoglobin.

Myoglobin or *muscle hemoglobin* (mol. wt. 16 700, 1 atom of iron) the pigment of muscle, resembles blood hemoglobin in its function. It acts as an oxygen reservoir within the muscle fiber which serves to tide the muscle over from one contraction to the next. It has a higher oxygen affinity than has blood hemoglobin and can combine with oxygen and dissociate from it with great rapidity (less than 1/100 second). Myoglobin starts to give up its oxygen at the instant that the muscle contracts. Its oxygen store is replenished during the resting state.

COMBINATIONS OF HEMOGLOBIN WITH GASES

Oxygen (see also chapter 32)

Hemoglobin combines with oxygen by virtue of the iron which it contains. The two elements combine according to the law of definite proportions, two atoms of oxygen uniting with each atom of iron in the hemoglobin complex. Thus

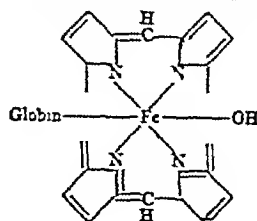


The combination of oxygen with hemoglobin is a most unstable one. When the oxygen pressure of the atmosphere in contact with hemoglobin is raised, oxygen is taken up but no true oxide is formed. The iron remains in the ferrous state. It is the globin which endows the iron-pigment part of hemoglobin with the unique property of forming a loose combination with oxygen; this property is lost once denaturation of the protein occurs. Hemoglobin *oxygenated* (not *oxidized*) in this way is known as *oxyhemoglobin*. The term *reduced hemoglobin* implies that the pigment has given up a part of its oxygen store.

The capacity of the blood for absorbing oxygen—the *oxygen capacity*, as it is called—is proportional to the hemoglobin concentration. The oxygen capacity of a gram of hemoglobin is 1.34 cc. So the oxygen capacity of 100 cc. of normal human blood (15 grams Hb) is $(15 \times 1.34 =) 20$ cc. The union of oxygen with hemoglobin will be dealt with more fully in the section on respiration.

Hemoglobin also combines with carbon dioxide (ch 32)

Methemoglobin (Ferrhemoglobin)



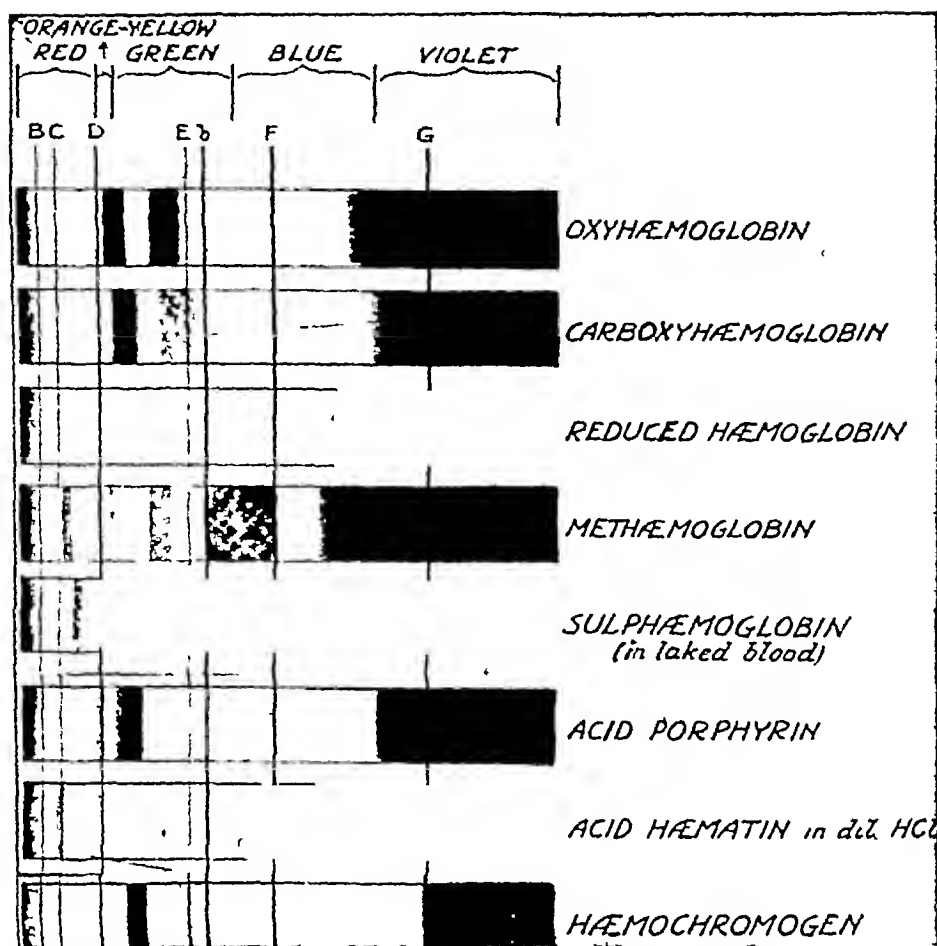
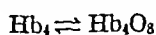


Fig 62 Absorption spectra of hemoglobin and some of its derivatives (after Peterson, Haines and Webster, *Legal Medicine and Toxicology*)

This is a true oxide. One atom of oxygen combines with one of iron, and the gas cannot be removed by exposing the blood to a vacuum, it can be removed only by chemical reagents. Methemoglobin is, therefore, a compound of oxidized heme (i.e., heme containing ferric iron) with native globin and thus differs chemically from carboxyhemoglobin which is ferric heme plus denatured globin, and from hemochromogen in which the globin is denatured but the iron is in the ferrous state. When potassium ferricyanide is added to oxyhemoglobin, as in the Haldane method of determining the oxygen content of a sample of blood (p 370), the two loosely bound atoms of oxygen are easily displaced but the iron is oxidized by the reagent and methemoglobin formed.⁷ In poisoning by

⁷The change is explained in the following way. The ferricyanide oxidizes the reduced hemoglobin which is in equilibrium with the oxyhemoglobin



The equilibrium is re-established by the passage of oxyhemoglobin into the reduced state, i.e., the oxyhemoglobin gives up its oxygen with the result that a loose compound of oxygen with hemoglobin is replaced by a true oxide of hemoglobin, i.e., methemoglobin

certain drugs, e.g., nitrites, chlorates, sulphates, acetanilid, bismuth subnitrate, nitrobenzine compounds, sulphanilamide, etc., the blood becomes dark in color due to the conversion of part of the hemoglobin to methemoglobin. The methemoglobin gives rise to a type of cyanosis (p 435) to which the term "toxic" is applied.⁸ The discoloration of the skin becomes evident when the methemoglobin amounts to about 3 grams per 100 cc of blood. Methylene blue when injected into the blood stream, as was practiced at one time to antidote cyanide poisoning (p 435), causes the formation of methemoglobin.⁹

Sulphemoglobin

Reduced hemoglobin combines with hydrogen sulphide to form sulphemoglobin which gives the

⁸According to Meulengracht and his associates, the cyanosis caused by acetanilid is due to the decomposition of the drug to dark colored derivatives of para-amido-phenol rather than to the formation of methemoglobin.

⁹Familial methemoglobinemia is a very rare congenital blood state in which Co-enzyme factor I (diphosphorase), required for maintaining the normal function of the hemoglobin, is deficient in the red cells. Cyanosis may be extreme.

blood a chocolate color. Except perhaps in extreme cases of intestinal putrefaction hydrogen sulphide is not absorbed in appreciable amounts. But it appears that certain drugs, notably acetanilid and phenacetin (drugs employed for the relief of headache), sensitize hemoglobin so that it combines more readily with hydrogen sulphide. Small quantities of the gas absorbed from the alimentary canal may then cause sulphhemoglobin to reach a relatively high concentration in the circulation and give a bluish or mauve tint to the skin. This so-called *enterogenous cyanosis* occurs when the abnormal compound amounts to from 3 to 5 grams per 100 cc. of blood. The presence in the blood of sulphhemoglobin or of methemoglobin is detected by spectroscopic examination (fig. 6.2).

Carboxy-hemoglobin and nitric oxide hemoglobin

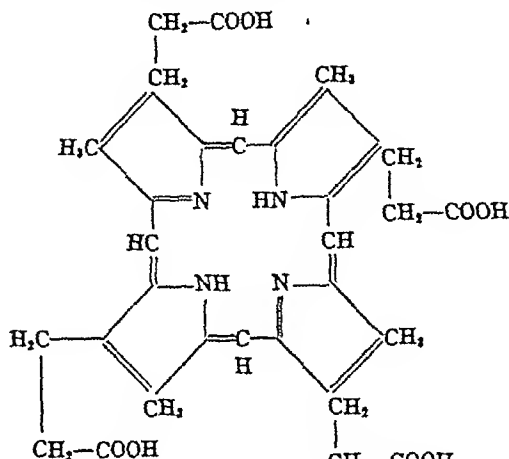
Carbon monoxide combines with hemoglobin in the same proportion as does oxygen. It competes successfully with the latter for hemoglobin and displaces it volume for volume to form carboxyhemoglobin. Unlike oxygen, however, it forms with hemoglobin a stable compound which can be disrupted only with the greatest difficulty. The much greater avidity (between 200 and 250 times) which hemoglobin shows for CO renders the gas so highly dangerous when inhaled in any considerable quantity (see p. 435). Nitric oxide gas also has a strong affinity for hemoglobin and forms a stable compound with it. The fumes given off by high explosives during their combustion contain large amounts of nitric oxide and the commonest way in which this poisoning occurs is through persons entering a closed space after an explosion, before the gas has cleared away.

PIGMENTS DERIVED FROM ANIMAL PORPHYRINS

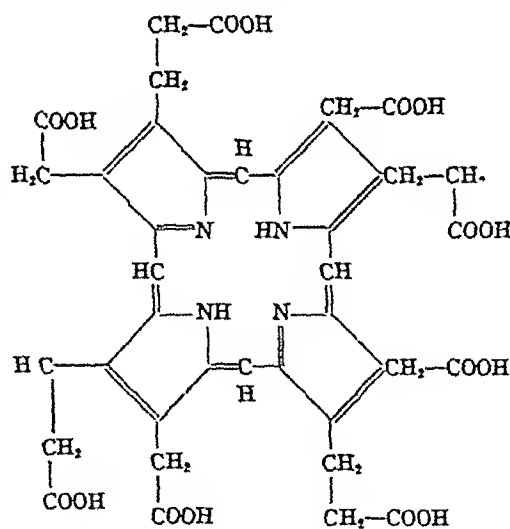
Hematoporphyrin, $C_{22}H_{28}O_8N_4(COOH)_2$, is an artificial derivative obtained by the action of strong mineral acids upon hemoglobin or upon hemin. It is closely related to protoporphyrin—the natural porphyrin of hemoglobin—but contains two more molecules of water in its vinyl groups.

Coproporphyrins and *uroporphyrins*. Coproporphyrins I and III can be obtained from the corresponding porphyrins (p. 55) by the replacement of two vinyl groups by propionyl groups, or by decarboxylation of uroporphyrin. Uroporphyrins I and III are formed by carboxylation of the methyl groups of the corresponding coproporphyrins. Coproporphyrin I is present in small amounts in normal urine and in larger amounts in the feces. Relatively enormous amounts of both types of coproporphyrin and of uroporphyrin are found in the

urine and feces in conditions of disordered porphyrin metabolism (see below). Coproporphyrin (chiefly Type III) is present normally in the bone marrow and the erythrocytes. It may be a precursor of protoporphyrin and a step in hemoglobin synthesis (p. 53). Turacin



Coproporphyrin I, $C_{24}H_{30}O_8N_4(COOH)_2$



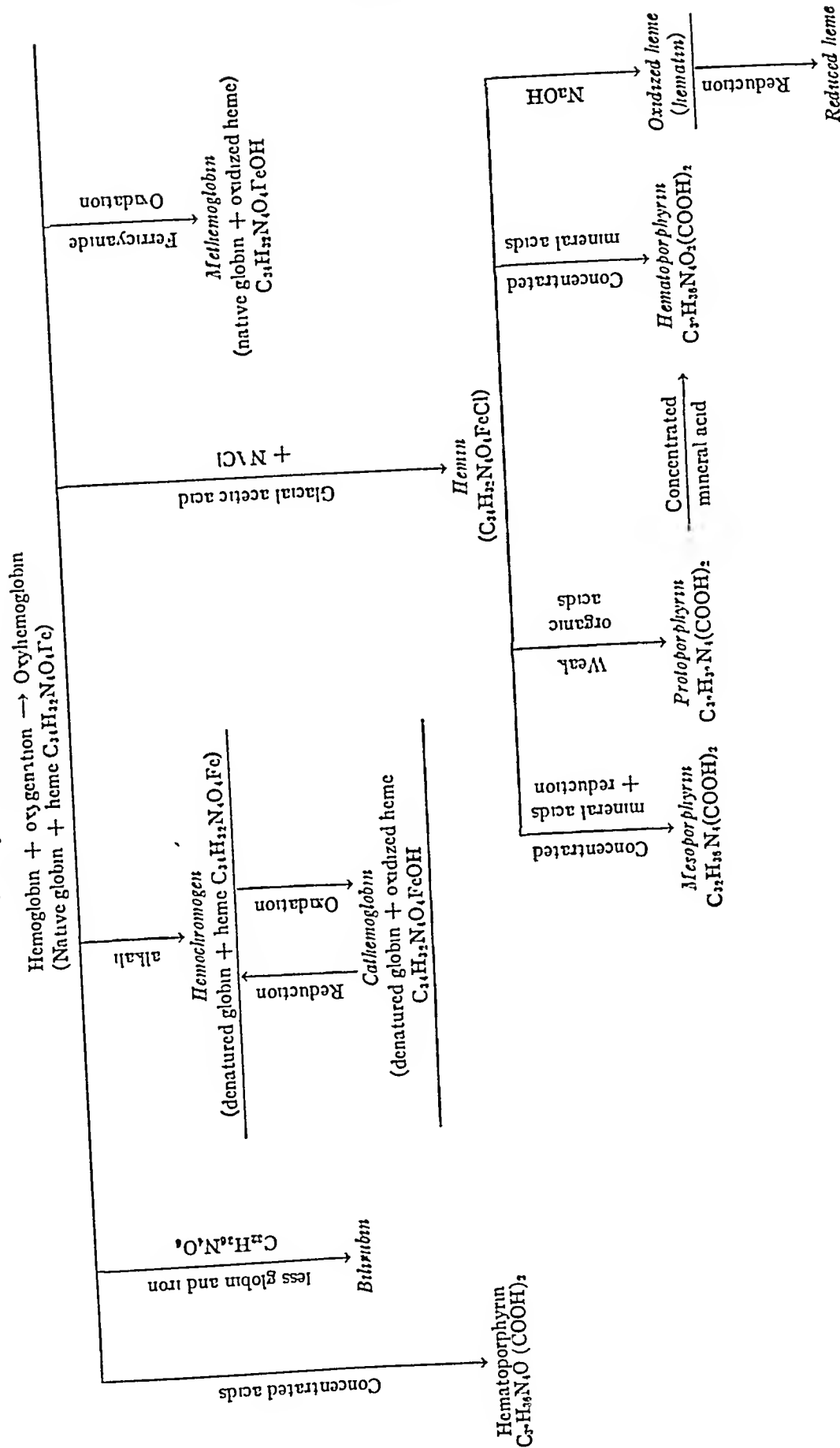
Uroporphyrin I

mentioned on p. 55 is the copper salt of uroporphyrin III. Porphyrins, especially hematoporphyrin, uroporphyrin I and coproporphyrin I have the curious property when injected into the blood stream, or when produced in disease, of sensitizing the skin to sunlight. Of the naturally occurring porphyrins the light sensitizing action is most pronounced with uroporphyrin I.

By the action of mild reducing agents the vinyl groups of protoporphyrin are replaced by ethyl groups to form mesoporphyrin.

Bilirubin, $C_{28}H_{36}N_4O_6$, is an iron free and globin free derivative of hemoglobin. The ring structure of the

Table of hemoglobin derivatives



porphyrin is broken at the alpha position, a straight chain resulting. This occurs *before* the iron and globin are split off so that no free protoporphyrin is produced normally in the circulation (see also p 537). Bilirubin is converted by reduction in the intestine to stercobilinogen, also called urobilinogen (p 539). Urobilin is formed by the oxidation of stercobilinogen, and is present in very small quantities in normal urine, but is not the pigment responsible for the color of urine (p 467).

Absorption spectra The various heme pigments and the compounds of hemoglobin when placed in the path of a beam of white light absorb waves of certain lengths but transmit the rest. That is, each possesses a characteristic absorption spectrum and it is by means of spectroscopy that they may be most readily detected. The different hemochromogens, reduced hemoglobin and oxyhemoglobin, CO and NO hemoglobins, hemato-porphyrin, coproporphyrin, cytochrome, urobilin, etc. have all their specific absorption bands. See fig 6.2 and table 10.

Porphyria Disorders of porphyrin metabolism in which abnormal porphyrins or large quantities of physiological porphyrins are formed and excreted are called porphyria. There are three types of this disease,

which have been termed *congenital*, *acute* and *chronic*. Actually, all three are congenital and familial.

Congenital porphyria is characterized by light sensitivity, which may result in blistering and, not uncommonly, necrosis of the skin upon exposure to sunlight, and the excretion of large quantities of coproporphyrin and uroporphyrin, type I, and sometimes small quantities of protoporphyrin. The normal ratio of type I and type III porphyrins is disturbed, the former type being produced in excessive amounts. The urine is stained a port wine color and the pigment is deposited in teeth, bones and skin. Liver extract and ascorbic acid seem to exert a favorable influence upon the disease. In *acute porphyria*, large amounts of porphyrin type III (protoporphyrin) are excreted. Light sensitivity does not occur, but there are abdominal, nervous or mental symptoms, e.g., severe cramps and vomiting, progressive paralysis often of an ascending type, and various forms of psychosis. In the *chronic* form of the disease, coproporphyrin type III and small amounts of coproporphyrin type I are excreted in the urine and feces. There is mild light sensitivity.

Porphyrimuria An increased excretion of porphyrins in the urine in other conditions than the porphyrias is termed porphyrimuria. Increased excretion occurs in various hepatic diseases, type I and to a less extent type III being found in the urine and especially in the feces. Type I coproporphyrin is present in excess in the excreta of patients suffering from several types of anemia—*pernicious anemia*, *hemolytic jaundice* and in *leukemia*. In the first of these it has been demonstrated in the megaloblasts, erythroblasts and reticulocytes. Increased excretion of coproporphyrin Type III occurs in *aplastic anemia*, *Hodgkin's disease*, *poliomyelitis*, *chronic alcoholism* and *lead poisoning*. In any condition associated with stimulation of the hemopoietic system porphyrin production is increased, and contrary to what might be expected it is type I rather than type III porphyrins which are produced and excreted in excess. The light sensitivity of *pellagra* may probably be explained by the excessive amounts of coproporphyrin (chiefly Type I) which are formed and excreted in this disease, but it is not unlikely that the porphyrimuria is merely secondary to hepatic involvement.

TABLE 10

Wave lengths (λ) at the points of maximum intensity of absorption bands of hemoglobin and its derivatives as well as some of the other heme compounds

COMPOUND	NUM BER OF BANDS	SITUATION OF ABSORPTION BANDS WAVE LENGTHS IN ANGSTROM UNITS				
Oxyhemoglobin	2	5769	5448	—	—	—
Reduced hemoglobin	1	5650	—	—	—	—
Carboxyhemoglobin	2	5709	5350	—	—	—
Methemoglobin	4	6300	5780	5400	5000	—
Sulphemoglobin	3	6180	5780	5400	—	—
Hemochromogen	2	5585	5275	—	—	—
	2	5580	5270	—	—	—
Reduced heme	2	6070	5820	—	—	—
Cytochrome	4	6046	5665	5502	5210	—
Protoporphyrin (in acid)	2	6000	5540	—	—	—
Urobilin	1	4900	—	—	—	—

CHAPTER 7

HEMOLYSIS AND SUSPENSION STABILITY OF THE BLOOD

HEMOLYSIS OR THE LAKING OF BLOOD

Under normal circumstances the plasma contains no appreciable quantity of hemoglobin. If normal blood be centrifuged the corpuscles are driven to the bottom of the tube while the supernatant plasma is clear but faintly straw-colored. Under certain conditions, however, changes may occur in the red cell which will allow the hemoglobin to escape into the surrounding fluid, which then becomes discolored. This is called *hemolysis* or *laking*, and may be carried out in a test tube by means of various agencies both physical and chemical. Certain biological substances, such as the toxins of bacteria, snake venoms, are intensely hemolytic. On the other hand, substances belonging to the class of immune substances or antibodies, and known specifically as *hemolysins*, are formed in the blood. These have the power to hemolyze foreign red cells. After the action of certain hemolytic agents the dim colorless outline of the red cells—shadow cells or “ghosts”—may be seen, they represent the incompletely destroyed framework of stroma¹ (See frontispiece.) Some of the means by which hemolysis may be induced will now be considered in greater detail.

(1) Hypotonic solutions

The membranes of plant and animal cells are semipermeable (p. 29). They allow the passage into the cell of water and various substances in solution, but offer a barrier to the entrance or egress of others. The red cell is no exception, it contains substances (p. 9) which cannot pass out, and is surrounded by a fluid (plasma) containing materials which cannot pass in. We have here then a minute and almost perfect osmometer (p. 29), and indeed much of our knowledge of osmotic phenomena has been gained from the study of the behavior of plant and animal cells when placed in solutions of different concentrations. In normal blood the plasma and the corpuscles are in

osmotic equilibrium, i.e., the fluids separated by the corpuscular membrane are isotonic. If, however, the dissolved substances in the plasma are diluted by the addition of distilled water, a flow of water into the corpuscles occurs. An osmotic pressure is developed within it which the cell membrane is unable to withstand. The cell swells and becomes globular, the membrane stretches and the hemoglobin is liberated.

The process is in reality rather more complicated than this. It is probable that the hemoglobin is not contained within the red cell merely as in a bladder, or even in a number of smaller compartments, but is closely bound in some way to the cell structure. One reason for this belief is that purely mechanical agencies will not liberate the pigment. The cell may be torn into the finest shreds, yet each minute particle still retains its hold upon the hemoglobin. The pigment, however, is soon released when the surrounding fluid is made hypotonic. This suggests that the cell structure consists of semipermeable partitions of almost infinite fineness.

The normal red cell offers a certain resistance to the disintegrating effect of hypotonic solutions. A slight lowering of the osmotic pressure of the surrounding fluid will not produce hemolysis. The normal percentage of salts in human plasma is approximately 0.94. Normal cells may be placed in a 0.6 per cent saline solution without being hemolyzed. The cell increases in volume, but the hemoglobin does not escape. The first trace of hemolysis appears when the saline concentration is about 0.42 per cent and the cell volume increased to about 145 per cent. At 0.35 per cent the cells are fully “laked”, i.e., hemolysis is complete, the cell volume is around 165 per cent of normal just before this occurs. The resistance which erythrocytes offer to the hemolytic action of hypotonic solutions is used in this way as an index of the fragility of the red cells.

In performing a fragility test a series of tubes is set up containing solutions of NaCl graded in strength from 0.9 to 0.20 per cent at intervals of 0.05 per cent above and below the range at which hemolysis is expected, and at intervals of 0.025 within that range. A sample of blood is introduced into the saline in each tube, and the

¹ When hemolysis is induced by certain reagents, e.g. linoleic acid, but the cell structure remains intact, the addition of electrolytes causes the reappearance of hemoglobin in the cells (“reversed hemolysis”). This phenomenon is due probably to shrinkage of the cells and the concentration of a residue of unliberated pigment and not to the return of hemoglobin to the cell.

one in which hemolysis commences and the one in which the process is complete are noted. A *mechanical fragility test* is sometimes employed, a sample of blood being placed in a flask with glass beads and rapidly rotated. The liberated hemoglobin is then measured. The mechanical fragility is increased in acholuric jaundice, certain other hemolytic anemias and in sickle-cell anemia. The degree of hemolysis may be determined by centrifuging the diluted blood sample and estimating colorimetrically the quantity of hemoglobin in the supernatant fluid.

In pernicious anemia the red cells have been found to be actually less fragile than normal whereas in other conditions, e.g., some forms of purpura (p. 122) and acholuric jaundice (ch. 9), their fragility is increased.

Cells which show reduced resistance to hypotonic saline may show a normal resistance to other hemolytic agents, e.g., lysolecithin (p. 65) and vice versa. Erythrocytes of spherical form, such as those characteristic of acholuric jaundice, show a lowered osmotic resistance mainly because the cells can not increase their volume by as great an increment as can normal cells without injury to the cell structure (see p. 9). The resistance to hemolysis by lysolecithin on the other hand is not influenced by the shape of the cells, the hemolytic process being chemical in nature.

The permeability of the membrane of the red cell is quite different from that of the capillary membrane which, as we have seen (p. 30), allows the passage of all crystalloid substances and to some extent of plasma protein, and is also freely permeable to hemoglobin. The membrane of human erythrocytes, on the other hand, is impermeable under physiological conditions to hemoglobin, the plasma proteins, and to Ca^{++} , K^+ , Mg^{++} and organic phosphate ions, but permits the passage of water, H^+ , NH_4^+ , and of Cl^- , HCO_3^- and PO_4^{--} . Potassium escapes freely from the cells under conditions which injure the cell membrane, the potassium loss is closely related to the escape of hemoglobin, i.e., to hemolysis. The human red-cell membrane is not absolutely impermeable to the sodium ion for even under physiological conditions and, therefore, in the absence of hemolysis minute amounts may pass from the plasma into the erythrocytes.² Since the red cell is impermeable

to potassium while sodium can cross the membrane, the selective permeability of the latter cannot be explained simply upon the theory that the cell membrane is a sieve-like structure whose "pores" are of such a size as to allow the smaller ions, but not the larger ones, to pass. The lipid-soluble theory is also unsatisfactory, for the inorganic anions are lipid-insoluble. The cell membrane is freely permeable to amino acids, urea and uric acid, so these substances under ordinary circumstances do not enter into the osmotic relationships between cells and plasma. Osmotic changes occur, however, when CO_2 enters the blood and diffuses into the cell (see p. 397).

The exchange of inorganic phosphate across the cell membrane is associated with the enzymatic synthesis and breakdown of organic phosphate esters within the erythrocyte. In the former process inorganic phosphate passes from the serum into the cell and in the reverse direction when organic phosphate compounds are broken down.

(2) Chemical substances

Ether, chloroform, benzene and alcohol act by dissolving the lipid constituents of the envelope and stroma of the cell. Other substances, e.g., *bile salts, acids and alkalis and saponin* cause hemolysis, but the manner in which they act is not altogether clear. Bile salts probably act by combining with the protein constituents, and saponin with the cholesterol. As a result of the chemical changes induced by either of these substances, destruction of the cell stroma—*stromatolysis*—occurs. Acids probably act by penetrating the cell and increasing the osmotic concentration within. Swelling and liberation of the hemoglobin occurs in a manner analogous to that of hypotonic solutions. The stroma is not as a rule destroyed. Alkalis, particularly ammonia, are powerfully hemolytic, as is also ammonium chloride. The NH_4 enters the cell and through the increase in osmotic pressure causes swelling and liberation of the hemoglobin. Stromatolysis accompanies hemolysis by alkali.

Certain chemical poisons such as carbonic acid, nitrobenzene, pyrogallol, ricin, arsenical preparations used in the treatment of syphilis, and many other substances are capable of causing red cell destruction.

(3) Substances of bacterial origin or formed in the animal body

(a) **SPECIFIC HEMOLYSINS** If blood is injected into the veins of an animal of another species, or as already mentioned (p. 42) into an individual of the same species but whose blood group is incompatible with the blood group to which the injected blood belongs, agglutination of the red cells of the donor occurs, and hemolysis follows as a secondary effect. But if a series of injections

² The results of various investigators in the past have differed rather widely with respect to the permeability of the red cell to the sodium and potassium ions, but the experiments of Kurnick who employed radio-active isotopes of Na and K give strong support to the statement made here.

of erythrocytes be injected over a period of days into the blood of another species the serum of the latter develops a substance which promptly destroys the foreign cells through a *primary* hemolytic effect quite independent of agglutination. This hemolytic reaction, which was first demonstrated by Bordet, is specific, that is to say, it is only the particular species of erythrocyte to which the animal has been sensitized by previous injections that is destroyed by the hemolytic substance. The latter on this account is known as a *specific hemolysin*. It belongs to the class of immune substances or antibodies. Bodies of similar nature cause the destruction of other foreign cells and are known as cytolytins and bacteriolysins. All are part of a general protective mechanism which the body is able to build up against the invasion of foreign cells. When referring to these and other immune reactions the substance which upon entering the body causes their development is referred to as the *antigen* (i.e., the foreign red cells in the case of hemolysins). The antibody itself (e.g., the hemolysin) is heat stable and is spoken of as the *amboceptor*. The latter which is specific requires for its action another body which is nonspecific, is present in all sera and is destroyed by heat. It is known as the *complement*. Three factors (antigen, amboceptor and complement) are therefore necessary for the hemolytic or bacteriolytic reaction. After the action of the hemolytic or bacteriolytic amboceptor has been annulled through destruction of the complement by heat, the reaction may be restored by the addition of any serum (i.e., by supplying fresh complement).

When serum which has developed a bacterial antibody is incubated with an emulsion of the particular bacteria which has served as antigen, a reaction occurs which "fixes" or binds the complement. The phenomenon is spoken of as *complement fixation*. These facts were applied by Wassermann to the diagnosis of syphilis, and by subsequent workers as a test for other diseases, e.g., tuberculosis. For example, the previously heated serum of a subject suspected to be suffering from syphilis is incubated with (a) an emulsion of syphilitic liver³ tubercle or typhoid bacilli respectively (the antigen) together with (b) complement furnished by normal guinea-pig's serum. If the suspected serum contains a specific antibody (amboceptor) for the antigen employed, the former will bind the complement to the latter, i.e., fixation of complement will occur. The foregoing is an account of a bacteriolytic system. An hemolytic system is employed to render the reaction visible. Washed sheep's corpuscles are added to the former system, together with the previously heated serum of a rabbit which has been sensitized to the latter cells by repeated injections. This serum supplies the hemolytic amboceptor but its complement has been destroyed. If the test is positive no hemolysis of the

corpuscles occurs, since the complement (non-specific) of the patient's serum has been already fixed by the bacteriolytic amboceptor and the hemolysin of the rabbit serum is therefore unable to exert its usual effect.

(b) **TOXIC SUBSTANCES OF BACTERIAL OR PARASITIC ORIGIN** *Endogenous* poisons of unknown origin. The toxins of bacteria responsible for many diseases, e.g. streptococcus, staphylococcus, tetanus bacillus and the organism of scarlet fever may cause a destruction of red cells, it may also occur in extensive burns. The more virulent types of other infectious fevers, e.g., smallpox, diphtheria, are also sometimes accompanied by intense hemolysis. When the hemolysis is of moderate degree but occurs over longer periods the hemoglobin is converted into bile-pigment. This, if formed in amounts greater than can be disposed of by the liver, undergoes partial retention in the plasma, which together with the solid tissues, especially of the skin and mucous membranes, becomes stained a yellowish tint—hemolytic jaundice (p. 80). In hemolytic states of long standing an iron-containing derivative of hemoglobin termed *hemosiderin* is frequently deposited in large amounts in the tissues, particularly of the liver and spleen (p. 79).

(c) **THE VENOMS OF CERTAIN POISONOUS SNAKES**, e.g., the cobra, and the poisons of various stinging insects and spiders cause a destruction to a greater or less degree of the red cells. Snake venom (cobra) has been shown to act indirectly. It contains a principle which has power to remove unsaturated fatty acids from the lecithin molecule. The resulting product which is called *lysolecithin*, is intensely hemolytic. Since lecithin is present both in erythrocytes and plasma and indeed in all cells, the entrance of snake venom into the body causes the production of this intensely hemolytic substance. Cephalin is acted upon by snake venom in a like manner with the production of a lysocephalin which has a similar hemolytic action.

(d) **HEMOLYSINS FROM NORMAL TISSUES**. A non-specific hemolysin has been extracted by simple means from healthy tissues and identified as *cis-vaccinic acid*, it probably exists in the free state in the tissues. Other non-specific hemolysins have been isolated from various tissues after incubation at 37° and extraction with alcohol. These are thought to be present bound to an inhibitory substance, but may be liberated in active form by disease or injury.

HEMOGLOBINURIA

When the hemolysis reaches such a degree that the hemoglobin cannot be converted into bilirubin as rapidly as it is liberated, as in severe malaria (black-water fever) *hemoglobinuria* occurs, that is, the pigment is passed in the urine, which is usually turned port wine color or a dark brown or even black, due to the action of the urinary acid in converting the pigment into acid hematin and methemoglobin. The concentration of hemoglobin must, as a rule, reach a level of about 0.13 gram per 100 cc. of blood before it appears in the urine.

³ As a matter of fact syphilitic liver has been found to be unnecessary, since lecithin and other materials for some unexplained reason will serve as antigens.

It should be remembered that hemoglobin once it has escaped from the erythrocytes is functionless. Not only is it unable to be retained within the capillaries on account of the relatively small size of its molecule but the environment of the plasma is unsuitable for its action.

The hemolysis in *blackwater fever* is not due to the destruction of erythrocytes by the malarial parasite, nor does the parasite affect the hemopoietic tissue in such a way that it produces cells which are unduly fragile. The cells are apparently not defective, but the work of Macgrath and his associates may provide an explanation of the high degree of hemolysis in this disease. They found that normal human tissues contain a hemolysin which is inhibited in its action by normal serum but not by the serum of a patient suffering from blackwater fever. It is suggested, therefore, that there is no special hemolytic agent in the blood in blackwater fever but rather an absence of a normal inhibitory substance.

Paroxysmal (cold) hemoglobinuria This form of hemoglobinuria occurs most usually upon exposure to cold. There may be fever, headache, abdominal pain, vomiting and transient jaundice. It is sometimes associated with Raynaud's phenomenon, in which condition spasmodic constrictions of the small vessels of the peripheral parts of the body occur, particularly after exposure to cold. The occasional association of the two conditions has suggested to some that they have a common cause, but direct evidence for this is lacking. It is an interesting observation that the blood of a subject of paroxysmal hemoglobinuria if cooled (to 5°C.) outside the body and subsequently warmed undergoes hemolysis (Donath phenomenon). There is apparently no defect, however, of the subject's corpuscles, they seem to be no less resistant than normal to hypotonic saline, the serum, on the other hand, has the power to hemolyze the cells of a normal person. The great majority of subjects are syphilitic. The hemoglobin, it seems, is liberated from the erythrocytes by the action of an endogenous hemolysin which in the presence of complement becomes fixed to the red-cells when the blood is chilled (cold phase). The presence of some thermolabile component of complement is required when the blood is again brought to ordinary body temperature (warm phase) to bring about hemolysis (see Siebens and associates). The phenomenon is quite distinct from cold agglutination (p. 50).

March hemoglobinuria Hemolysis may occur and hemoglobin appear in the urine even in healthy persons after strenuous muscular effort. In certain persons this tendency is exaggerated and hemoglobinuria may follow relatively mild muscular exercise. It is seen not infrequently in soldiers after long marches. The free pigment in the blood and urine of such cases is oxyhemoglobin, not myoglobin as might be expected. The fragility of the red cells is not increased and neither hemolysins nor autoagglutinins which might account for the hemolysis have been discovered.

In the third type of paroxysmal hemoglobinuria—the *nocturnal hemoglobinuria of Marchiafava*—hemoglobin or hemosiderin (p. 79) is passed almost continuously in the urine, but in greatest amounts at night. It is accompanied by a severe hemolytic anemia. The cause of the hemolysis is believed to be a peculiar susceptibility of the red cells to acid metabolites, their greater destruction at night being attributed to the accumulation of carbon dioxide during sleep. The abnormal cells are hemolysed when incubated in serum made slightly acid (Ham), or in any normal serum, but normal cells are not destroyed by the patient's serum. The fault in the patient's cells is thought to be some abnormality of the stroma protein.

THE SUSPENSION STABILITY OF THE BLOOD ERYTHROCYTE SEDIMENTATION RATE (E. S. R.)

The blood is a suspension of cells in a viscous fluid, the plasma. It is only the constant movement of the fluid that keeps the cells evenly distributed throughout. When the circulation comes to rest the cells at once commence to sink. Under ordinary circumstances the sedimentation of the cells in a sample of blood can progress to only a negligible extent, for it is soon circumvented by the clotting process which fixes them in a jelly-like matrix. If for any reason, the blood is delayed from clotting, sedimentation may continue until an upper layer of clear plasma becomes separated from the cells which have descended through the fluid. When clotting then ensues the blood consists of two strata, a thin yellowish or buff-colored layer of clotted plasma laid upon a much deeper red stratum of cells. When blood had clotted in this way the upper layer was known to the older physiologists as the "buffy coat" (see also p. 113). Ancient and medieval physicians carried out crude observations of the quantity of clear fluid which separated from blood upon standing as a means of diagnosis. In conformity with their humoral theories of disease, they believed it to be the "phlegm" which had separated from the other humors.

Biernaki (1891) was the first in modern times to draw attention to the increase in the rate of sedimentation of the blood in various pathological states. The subject has been studied in more recent times by Fahraeus and the *erythrocyte sedimentation rate* (E. S. R.) has come to be recognised as a useful diagnostic procedure. The sedimentation rate is measured by the depth in millimeters of clear plasma which is formed at the top of a vertical column of blood at the end of one hour. For the determination of the E. S. R. either

Westegren's or Wintrobe's method is usually employed. In Westegren's method, the blood for examination (about 15 ml) is diluted 4 parts to 1 of a 3.8 per cent solution of sodium citrate.⁴ It is then drawn into a graduated glass tube about 300 mm long, and having a bore of 2.45 mm.⁵ The upper end of the tube which is fixed in a strictly perpendicular position is left open, while the lower end is closed, usually by a removable rubber cap held in position by a spring.

The determinations should be made at a temperature of 20°C.

The sedimentation rates expressed as the height of supernatant plasma, in mm per hour for normal blood of men, women and infants, are given in the following table:

	<i>mm per hour</i>
Men	1-3
Women	4-7
Newborn children	0-5

In normal pregnancy and in certain pathological states, the sinking rate of the red cells is found to be very markedly increased, in other words, the suspension stability of the blood is reduced. The average figure during pregnancy is about 35 mm per hour. The rate is also increased during menstruation. The pathological states which show the most noteworthy increase in the rate are septicemia, 100 mm per hour, and pulmonary tuberculosis, 65 mm. Anemia (sickle-cell anemia and acholuric jaundice excepted), malignant tumors, inflammatory conditions of the female pelvic organs and many other conditions increase the rate moderately above the normal. *Reduction* in the sedimentation rate is rare; it occurs in allergic states, in peptone shock and in sickle-cell anemia and acholuric jaundice.

The physical changes in the blood which might cause this unusually rapid rate of red cell settling are of considerable interest and were investigated thoroughly by Fahraeus. In considering the sedimentation rate of particles suspended in a fluid when they are of a size comparable with that of

the red cells, four factors must be taken into account. These as applied to blood are:

(1) *Specific gravity* of the plasma as compared with that of the corpuscles. Corpuscles of high specific gravity would sink more quickly in normal plasma, and normal corpuscles more quickly in a plasma with a low specific gravity. In neither corpuscles nor plasma was any significant change of this nature found to explain an increase in the ESR.

(2) *Lowered viscosity* of the plasma is another factor which could cause an increase in rate of sinking, but no such change could be detected.

(3) *Increased size of the corpuscles* would increase their mass disproportionately to their surface and in consequence enhance their rate of sinking, but no significant alteration in size was found.

(4) *Clumping* together of cells of normal size would have the same effect as an increase in size of the individual cells for just as lumps of clay sink rapidly in water, while clay in the form of fine particles remains suspended almost indefinitely, so aggregation of the corpuscles would cause their more rapid sedimentation.

This last factor is the chief cause of the lowered stability of the blood suspension in the pathological conditions cited above. The roughness and granular appearance of the blood, due to the corpuscular aggregation is evident to the naked eye when the blood is spread in a film upon a slide (fig. 7.1). Normal blood, in marked contrast, forms a smooth homogeneous film. Under the microscope the crowding together of the cells in large masses is quite obvious. An increase in the fibrinogen and euglobulin fractions of the plasma is held responsible for the effect. These proteins act upon the corpuscles in some unknown way to make them adhere to one another and form clumps of agglutinated cells.⁶ It is probable that it is the relative proportions of the fibrinogen, globulin and albumin fractions of the plasma rather than their absolute concentrations which are of importance in determining the ESR. Thus a fall in the albumin concentration alone may have as great an accelerating effect as a rise in the other fractions. That the character of the plasma and not that of the cells is the principal determining factor is shown by the fact that if erythrocytes from blood with a high sedimentation rate (e.g., of pregnancy) are suspended in the plasma of blood having a low rate of sedimentation (e.g., of newborn) they settle at the slower rate, and conversely, erythrocytes

⁴This is isotonic with plasma and has the same specific gravity.

⁵Since the concentration of the blood in red cells influences the sinking rate, Walton recommends that their number be standardized to 5,000,000 per cu. mm by the addition or removal of plasma if the subject's blood is below or above this level. In Wintrobe's method his hematocrit tube is employed, the packed cell volume (p. 15) being determined if desired after measuring the ESR. In the final calculation of the latter, a factor is used to correct for any existing anemia.

⁶The red cells of normal blood show an incipient tendency to cling together in chains—the so called *rouleaux formation* (pseudo-agglutination).

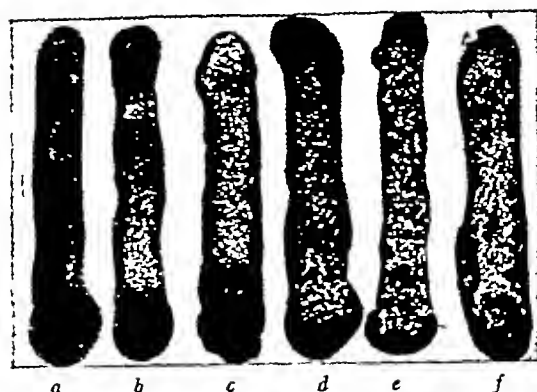


FIG 71 Showing the naked-eye appearances of specimens of blood spread upon glass slides. The specimens, left to right, are from, *a* healthy man (2), *b* healthy woman (7), *c* healthy pregnant woman (28), *d* man, appendicitis (40), *e* man, pneumonia (68), *f* sepsis (102). Note the especially granular appearance of specimens *e*, and *f* (After Fabraeus)

with a normal sedimentation rate in their own plasma settle rapidly in plasma from blood with a high ESR. See table 11.

Though the protein constitution of the plasma is usually the most important single factor affecting the sedimentation rate, several other factors exert an influence, e.g., (1) the *shape* of the erythrocytes, any tendency toward a spherical form by reducing the proclivity of the cells to cling together (rouleaux formation) will retard

sedimentation, (2) *temperature*, a rise in temperature above 20°C, as a rule, accelerates the ESR, a fall retards it, (3) the *lecithin-cholesterol ratio* of the plasma, the rate is retarded by an increase in the lecithin concentration and accelerated by cholesterol, (4) *red-cell concentration*, a high erythrocyte count retards the ESR, a low count accelerates it.

The non-specificity of the test is evident, nevertheless determinations of the sedimentation rate are of considerable value, especially, (1) in gauging the degree of activity of tuberculous processes, (2) as an aid in the differential diagnosis of certain gynecological lesions. Benign tumors of the pelvic organs cause no change in rate whereas, as already mentioned, malignant growths, inflammatory states and pregnancy cause a pronounced rise, (3) as an index of the extent and intensity in pyogenic infections, and (4) in estimating the activity of the inflammatory process in rheumatic fever.

TABLE 11

Showing the sedimentation rates of corpuscles from the same sample of blood suspended in different protein fractions of plasma (from Zoraya)

PROTEIN FRACTION	ESR.
Fibrinogen	41-61 mm.
Euglobulin	42-64 mm.
Pseudoglobulin	5-12 mm.
Albumin	3-6 mm.

CHAPTER 8

THE SPLEEN, THE LIFE OF THE RED CELL, T OF BLOOD, IRON METABOLIS

THE FUNCTIONS OF THE SPLEEN

The spleen serves three well-recognized purposes, namely, (1) the final disposal of the red blood cells, (2) the storage of blood, (3) actions possibly hormonal in nature upon the longevity, form and structure of the erythrocytes and upon the number of platelets and leukocytes, (4) the manufacture of lymphocytes in the lymphoid tissue composing the Malpighian corpuscles (p 71)

(1) THE RÔLE PLAYED BY THE SPLEEN IN THE DESTRUCTION OF THE BLOOD CELLS

In the pulp of the spleen are to be found relatively enormous mononuclear ameboid cells which have the power to engulf foreign particles of various sorts. They are known as *macrophages* and at times may be seen with fragments of erythrocytes or even whole corpuscles within their cytoplasm. These cells belong to the reticulo-endothelial system (p 105). In certain conditions in which a great destruction of red cells is a feature, immense numbers of these phagocytic cells may be seen loaded with erythrocyte fragments of various sizes. Sometimes merely a dust-like residue (*hemoconia*) containing hemoglobin is all that remains of the blood cell. The disposal of abnormal or over-mature erythrocytes by the spleen may not be a phagocytic process entirely, for it has been postulated that the cells in their slow progress through the splenic circulation are subjected to the action of a physiological hemolysin, possibly *lysolecithin* (p 65) or a substance allied to it.

Attempts to demonstrate the red-cell disposal function of the spleen by comparative estimates of the corpuscular contents of the arterial (incoming) and venous (outgoing) bloods have not, on the whole, been very successful. But Mann and his associates have been able, by spectroscopic examination of the arterial and venous bloods, to show a definite excess of bilirubin (iron-free pigment) in the blood of the splenic vein over that of the splenic artery. The bilirubin in the venous blood of rats, for example, was a greater amount

impregnated with a pigment, *hemoside*, globin of the disintegration occur in the liver tissues.

Though evidence from studies and from it is undoubtedly that red spleen, it is believed or effete and senile abnormal are disposed not thought to attack rather serves as a "ter-house" for the

(2) THE SPLEEN

This function was an accident, and the discovery might say, of researches conducted with another party which travel make observations with acclimatization sickness. Blood examinations were made upon board ship from a voyage in order that obtained with white altitudes later could reach tropical waters the hemoglobin concentration increased. Upon passing colder regions it was surmised that temperature variations mechanism that could destroy red cells at rapid changes in concentration. It is the extra blood hemoglobin climatic rise in temperature where it had been to the reservoir of

The spleen
(2), it may
there is a specimen

in temperature
the E.S.R.
ratio of
an increase
accelerated by
a high
S.R., a low count

in extent, never
sedimentation rate
(1) in gauging
processes, (2)
of certain
of the pelvic
whereas, as al
inflammation
a pronounced
and intensify in
in estimating the
process in rheumatic

When guinea pigs were permitted to breathe an atmosphere containing from 0.06 to 1 per cent carbon monoxide, it was found that the blood of the general circulation contained the expected amount of the inhaled gas but the blood of the spleen contained none until a considerable time had elapsed. There was a lag in some cases of 2 hours in the absorption of the gas by the blood of the spleen. When, after the splenic blood had absorbed the gas, the animals were brought into pure air, the carbon monoxide disappeared from the systemic circulation much sooner than from the blood of the spleen. These facts could only be accounted for on the assumption that blood had been held in the spleen out of the general circulation, i.e., in a sort of cul de sac. No difference in the gas content of the bloods in the two regions occurred, however, if strenuous exercise were carried out immediately prior to the inhalation of the gas. This suggested that during exercise the spleen contracted, and expelled the blood that had been held stagnant within its sinuses. Proof was gained by an ingenious experiment. Metal rods were fastened to the spleen at such points that changes in their position would indicate changes in splenic volume. The animal's abdomen was closed and time allowed for it to recover from the operation. X-ray examinations were then made before and after exercise, when it was found that a pronounced shrinkage of the organ occurred as a result of the exertion. In other experiments the spleen was drawn out of the abdomen and fixed to the abdominal wall which was then closed around the splenic pedicle. The animal recovers after this operation and remains in good health. Direct observations of changes in splenic volume could by this means be made from time to time under different experimental conditions. It has been estimated that in the cat the spleen is capable of expelling during very strenuous exercise a quantity of blood (plasma and cells) equal to one-sixth of the total blood volume, and that the number of red cells discharged might be as much as one-fourth of the body's total supply. The blood of the spleen has therefore a much higher corpuscular content than that of the general circulation. The concentration of the blood in the spleen occurs very rapidly—within a few minutes.

By the liberation of a large number of red cells during exercise the blood is enriched in hemoglobin, and its oxygen carrying capacity as a consequence increased. In this way the spleen serves

an *emergency function*. Its value in an emergency was shown when normal and splenectomized guinea pigs were exposed to an atmosphere containing a given percentage of carbon monoxide. The animals operated upon succumbed much sooner than did the normal ones, and a dose just sufficient to kill the former was not lethal for the latter. There was no difference, however, in the survival time when hydrocyanic acid was used, since this poison does not cause death through any effect upon the oxygen carrying capacity of the blood. In conditions such as muscular exercise, hemorrhage, carbon monoxide poisoning, etc., or under any circumstances in which the oxygen supply to the tissues falls below their requirements, contraction of the spleen and the expulsion of an extra quota of blood into the circulation occurs. The necessity automatically brings about its own relief, for lowered oxygen tension is the adequate stimulus for splenic contraction. The stimulus evidently acts not directly upon the spleen but upon the nerve centers, for there is no response to lowered oxygen tensions if the splenic nerves have first been cut.¹

Cannon and Izquierdo have shown that in cats excitement produces an increase of over 25 per cent in the red cell count, which they ascribe to a discharge from the splenic reservoir. Injections of adrenaline, pituitrin, acetylcholine and pilocarpine have a similar effect. Hargis and Mann have demonstrated in dogs that the slightest disturbance in the animal's surroundings, e.g., the banging of a door, or procedures causing minor discomfort to the animal induce reflex-contraction of the spleen. This is accompanied by increased blood flow through the splenic vein. The effect of emotional states upon splenic volume are shown in figure 8.1.

The spleen also undergoes spontaneous rhythmic changes in volume which as shown by Barcroft and Nisimaru are responsible for certain low undulations seen in blood pressure tracings. These were first observed by Roy in 1881. They have a duration of from 25 to 50 seconds (fig. 8.2). When the splenic and blood pressure tracings are synchronized, the peaks of the blood pressure waves coin-

¹It was observed by Barcroft and his colleagues that the splenic volume progressively diminishes from about the middle of pregnancy to term. Blood is transferred apparently to the uterine vessels for the nourishment of the fetuses. After the birth of the young the maternal spleen rapidly regains its usual size. In animals in which the spleen has been denervated, pregnancy does not cause a reduction in splenic volume.

cide with the troughs of the splenic fluctuations. The waves are more pronounced after denervation of the spleen, they disappear after clamping the splenic artery. The fluctuations in splenic volume were believed to be due to the rhythmical contractions of the smooth muscle of the spleen, but it has been shown by Mertens and by Grindlay and his associates that they are caused by periodic changes in the splenic blood flow, and originate in the arterial or arteriolar musculature. The rhythmical changes in the blood flow were recorded by means of the thermostromuhr (p. 176) from the splenic artery and vein while the periodicity of the waves recorded from the two vessels was identical, the venous fluctuations lagged behind those of the arterial blood flow and splenic volume by an interval of about 5 seconds.

The structure of the spleen is suited most admirably for its reservoir function. In its relaxed condition in the living body it is a large organ—3 to 5 times larger than post-mortem examination would lead one to suspect, for the throes of death cause it to contract to minimal proportions. The blood is delivered into the substance of the spleen by fine arterial vessels. Through these the blood floods the splenic pulp and percolates between its cells. It is collected again into venous sinuses whose walls are formed of flattened cells possessing phagocytic proclivities of the highest order (reticulo-endothelial elements). The blood passes freely into the sinuses through numerous perforations in their walls. A certain proportion of the splenic arterioles open directly into the sinuses. The latter are drained by small veins with intact walls which join with similar radicles to form larger venous trunks. The splenic circulation is therefore an open one and a large part of the blood in its passage from

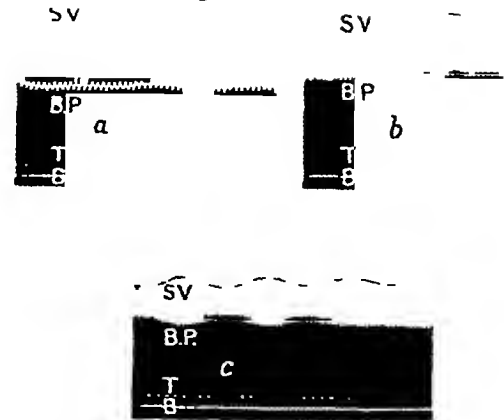


FIG. 82 SV = splenic volume BP = general arterial blood pressure T = time, 5 sec B = baseline at 40 mm Hg. Spleen in plethysmograph, a, before clamping splenic vessels, b, after clamping vessels, c, after removal of clamp (after Barcroft and Nisumaru).

the arterial to the venous side traverses the morass of splenic pulp. The flow at ordinary times is very sluggish, and within the venous sinuses the blood may be almost stagnant. Nevertheless within a relatively short time all the red cells in the body are brought into direct relationship with the splenic tissue to pass inspection by the macrophages lurking in the pulp and lining the sinuses. To these cells the infirm, senile or dying erythrocytes fall a prey.

The capsule of the spleen is provided with *unstriated* muscle fibers which penetrate its substance and are continued along the fibrous trabeculae into the depths of the organ. The distribution of the muscle in this way explains how the spleen is enabled so quickly to alter its volume.

Dotted throughout the spleen, like islands, and surrounded by the pulp are lighter areas of lymphoid tissue. These are the *Malpighian corpuscles*. Many of them, especially in infancy, show a pale central area known as the "germ center." The Malpighian corpuscle (fig. 8.3) is pierced by a small artery, it is analogous to similar areas in lymph nodes, and serves the same function, namely the manufacture of lymphocytes. The development of these cells is considered in chapter 11.

(3) THE HEMOPOIETIC FUNCTION OF THE SPLEEN IN THE EMBRYO AND ITS POSSIBLE ACTION UPON THE BLOOD CELLS IN POSTNATAL LIFE

In the embryo, the spleen in common with the bone-marrow and liver is active in the production of erythrocytes and granulocytes as well as of lymphocytes. But erythropoiesis and granulopoiesis in the spleen normally cease at birth. Yet, many observations point to the spleen as exerting an influence upon the form and structure of the erythrocytes, while they move slowly through the

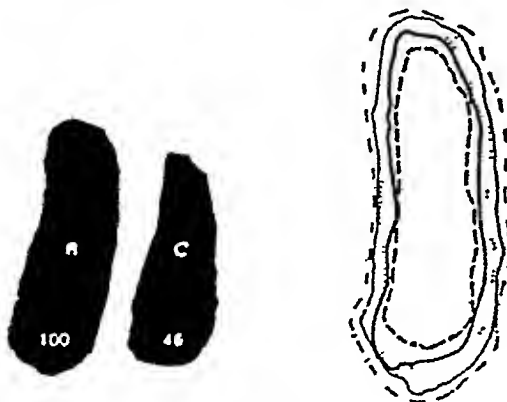


Fig. 81 Changes in volume of spleen as a result of emotional excitement (after Barcroft). Sketch on left, R, rest, C, dog sees cat. The numbers represent the relative sizes of the dog's spleen. Sketch on right, — — — rest, ——— smells cat, ———— hears cat, ———— sees cat, ———— chases cat.

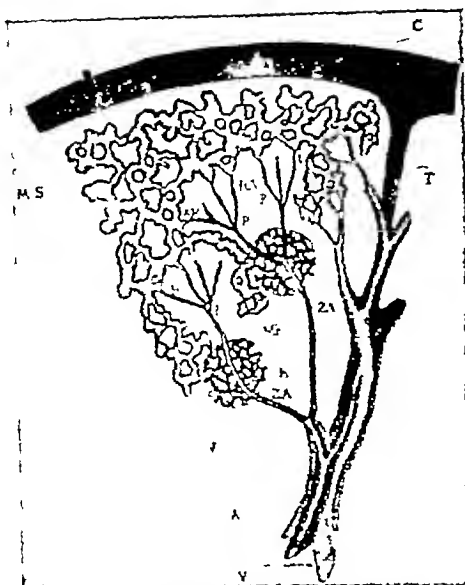


FIG 8.3 Diagram of the human spleen (A) artery and vein (V) in a trabecula (T) of the capsule (C), ZA, central artery of the corpuscle of Malpighi (VK), P, small arteries, HA, arterioles with a sheath, AK, arterial capillaries which terminate in the sinuses (I) or in the meshes of the reticulum (Z), MS, venous sinuses, W, white pulp (from Cajal after Szymonowicz)

splenic vessels or during their development in the bone-marrow. Though, as mentioned on page 69, the spleen destroys abnormal cells (e.g., the spherocytes of chronic congenital hemolytic anemia, p 80) and normal cells after they have served their purpose, it appears that a normal splenic function is to prolong the life of the erythrocyte. After splenectomy, the life span of the cell is definitely shortened. Such a decrease in the longevity of the red-cell is not, however, due to an increase in its fragility, for erythrocytes in the blood of the spleen actually appear to be more fragile than those of the general circulation. The red-cells are also thinner than normal in splenectomized animals and show nuclear changes from which it has been inferred that thickening of the cell and removal of the nucleus are processes within the sphere of splenic influence. Such observations suggest, though rather vaguely, an endocrine influence by the spleen upon the bone-marrow. But attempts to obtain a hormone having an effect upon hemopoiesis have not yielded any conclusive result.

The spleen is concerned undoubtedly in the life history of the platelets. In normal animals and in patients suffering from thrombocytopenic purpura (p 122) splenectomy increases the number of

circulating platelets, which implies that the spleen normally has a destructive action upon the platelets which may become enhanced under certain conditions.² After splenectomy animals become more susceptible to certain parasitic and bacterial infections, but there is no proof that this lowered resistance to such diseases is due to the removal of any specific effect, and not simply to the large population of phagocytic cells in the spleen as well as to its rich supply of lymphoid tissue—a source of antibodies.

ENLARGEMENT OF SPLEEN (SPLENO-MEGALY)

Enlargement of the spleen is associated with a large number of pathological states, only a few will be mentioned.

In *thrombocytopenic purpura* (p 122) the spleen is enlarged and its removal is followed by a rise in the platelet count and amelioration of the symptoms. Also, substances such as anti platelet serum or diphtheria toxin, which when injected cause platelet destruction, are rendered much less effective if the spleen has been removed. Splenectomy frequently brings about a cure in *hemolytic jaundice* (acholuric jaundice), p 80. The spleen may be considerably enlarged in *pernicious anemia* but splenectomy exerts no effect upon the course of the disease. (It has been mentioned that the spleen does not destroy healthy erythrocytes and the fragility of the red cells is actually decreased in pernicious anemia.) In *splenic anemia* (congestive splenomegaly, Banti's disease), the spleen is tremendously enlarged, the liver becomes cirrhotic. There are portal hypertension and repeated hemorrhages. Splenectomy has been practiced in this disease with variable success. Relief of the portal hypertension and splenic congestion by anastomosis of the portal vein to the inferior vena cava is advocated by Whipple. Splenic enlargement is a feature of many other abnormal conditions, such as malaria, Hodgkin's disease, leukemia, etc., but in these benefit does not follow its removal. The organ is also enlarged sometimes enormously in *polycythemia*, but splenectomy fails to cure the condition, and indeed is attended by grave risks to life (Moynihan).

Gaucher's disease is an interesting though rare condition which, commencing usually in childhood, is associated with a colossal enlargement of the spleen. The enlargement is due to hypertrophy and hyperplasia of reticulo-endothelial elements. Masses of very large

² A curious and unexplained action of the spleen has been discovered by Cullum and Simpson, who found that toxic doses of the antithyroid compound, thiouracil (2 thio-5-methyl uracil) causes the death of normal rabbits from pulmonary edema, but not of splenectomized animals. Moreover, if a splenectomized animal is injected with an aqueous-acid extract of hog or beef spleen, its susceptibility to the action of the drug in causing pulmonary edema is restored. Spleenic extracts alone produce no such effect.

vesicular cells (Gaucher's cells) filled with a cerebroside called *kerasin* are seen in the lymphoid tissue and venous sinuses of the spleen. Hyperplasia of reticulo-endothelial elements also occurs in other situations, e.g., bone marrow and liver. The disease is due apparently to some disorder of lipid metabolism. Splenectomy is the only effective treatment. *Niemann-Pick's disease* is a somewhat similar condition affecting the reticulo-endothelial system and lipid metabolism. Characteristic cells, known from their appearance as "foam cells", are present in large numbers, they are loaded with sphingomyelin. Accumulations of this phospholipid are found also in the liver and brain. Subjects of this and the preceding disease are usually children of the Jewish race. *Von Jaksch's disease* occurs in infants and is characterized by splenomegaly, anemia and an increase in the number of white blood cells. Great enlargement of the spleen also occurs in *glycogen storage (Gierke's) disease*. The chief characteristics are hypoglycemia, with ketosis, a slight rise only in the blood sugar after an injection of adrenaline (ch 59), and extensive deposits of glycogen in the spleen, liver, heart and other organs.

From its position in the portal circulation the spleen is also very susceptible to enlargement, either as a result of mechanical obstruction to the veins or to high venous pressure resulting from cardiac or hepatic disease.

THE LIFE OF THE RED CELL

From the amount of bile pigment which is excreted daily by the liver the conclusion must be drawn that a very large amount of hemoglobin (since this is the sole or at least the main source of bile pigment (p 537)) is liberated from disintegrated red cells in 24 hours.

There are three possible ways in which erythrocytes normally might be destroyed in the body: (1) By the macrophages of the spleen, (2) by the action of a hemolytic substance in the blood, (3) through simple wear and tear and disintegration in the blood stream. There is no evidence that hemolysis occurs to any significant extent in normal blood and the phagocytic cells of the normal spleen seem quite inadequate to account for the wholesale destruction of cells which evidently must be going on continually in the body. The work of Rous indicates that the erythrocytes to a very large extent undergo disintegration in the blood stream as a result of the stresses and strains to which they are incessantly subjected during their passage through the vessels. When it is considered how delicate is the structure of the red-cell and to what violent treatment it is exposed during its lifetime, this wastage is not surprising. The cells are flung from the heart into the arteries at high velocity. In their

voyage around the circulation they are exposed to jostlings and innumerable collisions with one another and with the arterial walls. At times they are forced through channels which are too narrow to permit their passage without marked distortions of their shape, or they may be caught in a fork at the branching of a vessel and become "saddle-bagged" over it. Their membranes are almost continually undergoing alterations in tension as a result of osmotic changes. At last, becoming older, they are unable to withstand these abuses and undergo fragmentation. Fragments of different shapes and varying in size from that of a half or a quarter of the whole cell, to mere dust-like remnants containing hemoglobin (hemoconia) may be found in the circulating blood, in the spleen and to a limited extent occasionally in other tissues.

From determinations of bile pigment excretion it has been estimated that in health between 7 and 10 million cells are destroyed in this way every second—and of course the same number must be formed afresh by the blood-forming tissue. The loss of hemoglobin is between 16 and 24 grams daily. The number of red cells and hemoglobin concentration in the circulation at any moment represents the balance struck between blood wastage and blood formation by the bone marrow.

Many attempts have been made to determine the life-span of the erythrocyte. Several methods have been employed, e.g., stimulation of the discharge of reticulocytes from the marrow and determining the time of their maturation, measuring the bile pigment excretion, and in the earliest attempts even such an unphysiological procedure as the injection of red-cells of the frog (recognizable by being nucleated) into the circulation of mammals. The simplest and most commonly employed method is some form of the *selective or differential agglutination technique*, introduced in 1918 by Ashley. This method consists in transfusing compatible red-cells and later examining the recipient's blood from time to time for the presence of the foreign corpuscles, which are counted in a hemocytometer. The foreign cells are distinguished from the transfused person's own cells by means of the serum of another belonging to a group with which the recipient's cells, but not the foreign cells, are incompatible³ (p 42). That is, the recipient's

³ Thus if corpuscles of Group O are transfused into a person belonging to Group AB then, when the recipient's blood is mixed outside the body with Group B (anti-A) or Group A (anti-B) serum, the recipient's corpuscles but not the foreign cells will be agglutinated.

cells are agglutinated but not the foreign (donor's) cells. Using this method, it has been found that the life of the average red cell is about 125 days.

Landsteiner, Levine and James have employed the M and N groups in a similar way, injecting M cells into an N recipient, or vice versa, and using anti-M or anti-N serum to agglutinate the *transfused* cells, leaving the recipient's cells unagglutinated. A more recent method is that which employs the isotope of nitrogen (N^{15}) to tag the red-cell in the bone marrow. This is done by feeding glycine into which the isotope has been incorporated, this amino-acid is utilized in the synthesis of protoporphyrin. It is necessary to the method that the isotope should not leave the cell and be utilized again for the manufacture of hemoglobin, but should persist throughout the life of the cell. This has been shown to be so. The life span of the red cell as determined in man by this method is 127 days. Whipple and Hawkins, by determinations of the bile pigment excretion, obtained a closely similar figure in dogs, namely 133 days.

It must be remembered that cells of blood transfused for therapeutic purposes may have a much shorter life, since the body into which they are injected may be in a state of health deleterious to their existence. Ashby has found that in pernicious anemia the average life of transfused cells was 110 days, in aplastic anemia 41 days, in malignant cases 52 days, while in transfusions given post-operatively to ordinary surgical cases, uncomplicated by severe anemia or malignancy, the survival time of the transfused cells averaged 124 days.

THE REGENERATION OF BLOOD

THE MATERIALS NECESSARY FOR ERYTHROCYTE FORMATION

(a) *The red cell stroma*. It is doubtful whether the materials required for the construction of the framework of the cell, e.g., nucleoprotein, globulin, lecithin and cholesterol are ever lacking. The body possesses large supplies of these materials and an ordinary diet contains them in adequate amounts. In certain anemias, however, e.g., pernicious anemia and Cooley's anemia, the failure of the body to assemble these materials for the manufacture of cell stroma may possibly be a fundamental causative factor.

(b) *Hemoglobin* is added to the red cell only after the cell's development has progressed to a certain stage (p. 109). Synthesis is believed to take place in

the nucleus of the cells in the bone marrow. The complete history of hemoglobin in the body has yet to be written, but it is now established that the body can synthesize pyrrole groups from simpler and readily available compounds, possibly, as suggested by Hans Fischer, from aceto acetic acid, glycine,⁴ proline, oxyproline and tryptophane have been proposed by others as possible building stones. Given the pyrrol group, synthesis of protoporphyrin is readily effected (ch. 6). An interesting experiment in this connection is described by Whipple and his associates. When hemoglobin was given intravenously to anemic dogs, the animals' hemoglobin was increased by an amount equal to that injected. At the same time the excretion of bile pigment was increased by a corresponding amount. This paradoxical result is interpreted in the following way. The pyrrole of the injected hemoglobin is excreted as bilirubin, while the globin part of the molecule is utilized for the production of new hemoglobin, the pyrrole groups of which must therefore be derived from some other source—food or body tissue—and synthesized to protoporphyrin. It is possible that food porphyrins differing from that in hemoglobin can be broken down into their constituent pyrrole groups which are then rebuilt into protoporphyrin. The supposed value of spinach and other green vegetables as hemoglobin builders has suggested such a process. Yet according to Whipple and his associates chlorophyll is not utilized by the dog for hemoglobin synthesis.⁵ It has already been pointed out (p. 56) that heme (porphyrin + iron) is a universal material and is present in the great majority of food stuffs. So here it might be supposed was a source of an almost unlimited supply of the necessary pigment element. Yet heme cannot be split by the digestive secretions and it is generally agreed that iron so combined can not be utilized for hemoglobin synthesis. Globin, of course, is required to complete the hemoglobin molecule. Whipple and his associates have found that this protein is well utilized for hemoglobin synthesis, yielding for each

⁴ The most convincing evidence for this amino-acid being a precursor of pyrrol has been obtained by Shemin and Rittenberg, they fed glycine to rats and to men, after incorporating isotopic nitrogen into its molecule, and found it in the newly formed protoporphyrin.

⁵ Hughes and Latner found that whereas large doses of chlorophyll do not favor hemoglobin regeneration, small quantities (15 mg.) are definitely stimulating. The larger doses they believe exert a toxic action (possibly due to the magnesium content of the pigment) upon the bone marrow.

100 grams fed to anemic dogs from 30 to 40 grams of blood pigment Hemoglobin or globin, or a digest of the latter when given intravenously to anemic animals, forms hemoglobin almost gram for gram. They found that as compared with the porphyrin, part of the molecule globin was of much greater importance for hemoglobin regeneration, this protein apparently is a limiting factor in hemoglobin synthesis. Upon a diet low in protein but adequate in iron, hemoglobin regeneration was minimal. The amino acids necessary for the construction of hemoglobin are present in milk, meat, and other sources of first class protein. Globin can also be synthesized to a limited extent apparently from endogenous sources. This function is probably situated in the liver. Certain amino-acids, especially proline and threonine, were found to increase hemoglobin regeneration after hemorrhage, which suggests that they are used for the synthesis of globin. Histidine which constitutes 8 per cent of the globin molecule is, contrary to expectation, less effective.

It has been stated in chapter 6 that both coproporphyrin Type III and protoporphyrin are present in the bone marrow and the erythrocytes. The coproporphyrin might be derived from protoporphyrin or be merely a by-product of hemoglobin synthesis. It seems more likely, however, that coproporphyrin is a precursor rather than a derivative of protoporphyrin and thus constitutes a step in hemoglobin synthesis. The evidence for this is that (1) in pernicious anemia the protoporphyrin of the erythrocytes is closely correlated with the reticulocyte percentage, and (2) when the disease is in relapse the protoporphyrin is very low and coproporphyrin cannot be demonstrated, but after treatment with vitamin B₁₂ the coproporphyrin increases rapidly and in advance of the protoporphyrin.

Whipple, Hooper and Rabschewitz carried out a series of experiments upon animals made anemic through repeated bleedings, and tested their power to regenerate hemoglobin when fed upon various diets. Meats were found to be the most potent for this purpose. Carbohydrates in the form of bread and sugar were found to be ineffective. In fact they had actually a definitely depressing effect upon the hemoglobin repair process, for animals regenerated their blood more rapidly when starved than when fed upon a bread and sugar diet. In explanation of this fact Whipple suggested that the starved animal drew upon its tissues to supply the basic elements

for hemoglobin synthesis, whereas carbohydrate food on account of its well-known protein-sparing effect (p. 639) prevented the tissues from being utilized in this way. Infection or very severe liver damage markedly depresses hemoglobin regeneration in anemic dogs. The depressing effect upon regeneration of hemoglobin which is seen in the bile and Eck fistula animals is apparently due to interference with liver function (defective protein synthesis) and to the reduction in iron absorption.

These workers found that, of all protein foods, liver was by far the most effective for blood regeneration. Next in order came kidney and chicken gizzard. Milk had little regenerating effect. Table 12 shows the comparative values of the various articles of diet. A bread mixture consisting of potato and wheat flour, bran, sugar and the necessary salts and vitamins was used as the basal diet. This was practically inert so far as the regeneration of hemoglobin was concerned. The article to be tested was added to this basal diet.

The animals (dogs) were rendered anemic by three or four successive bleedings until the hemoglobin had been reduced to 30 per cent of the normal. The item of food to be tested was then added to the basal diet and the animal bled from time to time in order to maintain the hemoglobin at the original level of 30 per cent. The

TABLE 12
Hemoglobin production influenced by diet

DIET, GRAMS DAILY	HEMOGLOBULIN PRODUCTION (TWO-WEEK FEEDING PERIOD)
	grams
Bread 400	3
Milk 450, Bread 400	3
Cream 100, Bread 400	10
Butter 100, Bread 350	15
Asparagus 200, Bread 300	9
Spinach 200, Bread 300	15
Raspberries 200, Bread 300	5
Raisins 200, Bread 300	25
Apricots 200, Bread 300	48
Eggs 150, Bread 300	45
Whole fish 250, Bread 300	13
Beef muscle 250, Bread 300	17
Pig muscle 250, Bread 300	30
Chicken gizzard 250, Bread 200	80
Kidney 250, Bread 300	70
Chicken liver 250, Bread 300	80
Beef liver 300, Bread 300	80
Beef liver 450	95

amount of blood removed expressed in grams of hemoglobin gave a direct measure of the amount of pigment regenerated in a given time.

IRON METABOLISM

Being an essential constituent of the hemoglobin molecule, this mineral must be available in adequate amounts in order that normal blood regeneration shall occur. A diet deficient in iron leads to a certain type of anemia (p 81). Iron provides the keystone for hemoglobin construction, unless it is supplied in appropriate amounts the maturation of the red cells is retarded, and the numbers discharged from the bone marrow into the general circulation reduced.

Absorption, storage and excretion of iron

Iron is absorbed to some extent throughout the entire intestinal tract, but in by far the greatest amount from the upper part of the small intestine. The absorption is by way of the blood. After absorption the element is stored in the intestinal mucosa, the liver and to a less extent in the spleen and kidney. Liver iron is readily increased by iron feeding or injection. Only minute quantities of iron are detectable in the plasma (0.1 to 0.3 mg per 100 cc.) under ordinary circumstances, the great proportion (60 per cent) of the iron of the blood being present in the red cells. A smaller quantity (22 per cent) is present in the muscle (myoglobin, cytochrome, etc.), the remainder is stored in liver, kidney, and various other tissues as *ferritin* and *hemosiderin*. Whole blood contains from 45 to 50 mg per cent and the total quantity in the adult human body is between 4 and 5 grams. See table 13. Iron is present in the blood in three forms: (a) plasma iron, already mentioned and

which represents mainly iron in transit from the intestinal tract to the depots, (b) iron combined with hemoglobin, which accounts for from 92 to 98 per cent of the total, and (c) from 2 to 8 per cent which is liberated by the action of mineral acids and called for this reason "easily split off" iron. This fraction is attached to the erythrocyte, and is derived apparently from an intermediary compound formed in the breakdown or synthesis of hemoglobin. This has been called *verdohemoglobin* (Lemberg and associates) or *pseudohemoglobin* (Barkan and Walker) (p 537).

It is commonly stated that iron is excreted almost exclusively through the intestinal wall, mainly through the wall of the colon. Evidence for this statement is lacking. If it were true one would expect an increase in fecal iron in conditions associated with the excessive breakdown of red cells, which is not the case. Also when iron is injected into rats almost all can be recovered from the tissues, a small fraction only appearing in the feces and urine, and Hahn and his associates found that, when radioactive iron was injected intravenously into dogs depleted of their iron stores, from 92 to 98 per cent was retained, the remainder appeared in the urine and feces within the first few days, excretion then fell to zero. Only very minute amounts of iron are found in human urine and bile, that appearing in the feces is mainly food iron which has not been absorbed. The view of McCance and Widdowson that the iron stores of the body are regulated not by excretion but through the control of absorption is now generally accepted. Iron is, therefore, very largely a "one way" element. That which is absorbed is held by the body with great avidity and the iron liberated from hemoglobin is used again for the manufacture of new hemoglobin, the pigment part of the molecule alone being excreted in the bile.

Several factors influence iron absorption, e.g., the acidity of the gastric juice, but especially the state of the *reserve stores of iron* in the body. Anemic dogs, for example, absorb a much greater proportion of administered iron than do normal animals, and Balfour and his associates found that from 2 to 10 times the usual amount of radioactive iron⁵ was absorbed in the later months of preg-

⁵ Radioactive iron has proved a valuable tool in studies of iron metabolism since it can be readily traced in the body tissues and fluids. It is prepared by bombardment of Fe^{56} isotope with deutron particles and detected in blood and other tissues by means of a Geiger counter.

TABLE 13

The distribution of iron in the body of a dog weighing 20 kilograms (Hahn)

	mg	per cent total body iron
Blood hemoglobin iron	900	57
Muscle hemoglobin iron	110	7
Total hemoglobin iron	1010	64
Parenchyma iron (muscle and other tissues)	240	16
Available visceral storage (liver, spleen, and marrow)	225	15
Available iron of other tissues (estimated)	75±	5±
Total iron	1550	100

nancy when the iron stores are low as a result of the demands of the fetus

When iron is fed in large amounts, that above the requirement is not absorbed. Even when the requirement is great and single large doses are fed, no more than a fraction enters the blood stream. Hahn has calculated that of the usual therapeutic dose of some 325 mg, only about 16 mg of the element is absorbed, which is sufficient for the manufacture of only about 5 grams of hemoglobin. This means that in anemia when the body may be required to synthesize some 300 grams, say, of the blood pigment several weeks of iron administration must elapse before the hemoglobin concentration will be restored to normal. That it is the depletion of the iron stores and not anemia itself which stimulates iron absorption is indicated by an experiment described by Hahn. When a well-nourished dog was bled suddenly by 60 per cent of its blood volume and then fed iron, no more was absorbed than before the hemorrhage, whereas after a week had elapsed (i.e., after depletion of the iron stores might be expected to have occurred) absorption increased many times over the control period.

The absorption of iron is favored by other subsidiary factors, e.g., chlorophyll, vitamin C, calcium, and bile pigments, but is retarded by alkalis, phytates, and phosphates, and by the excessive secretion of mucus. Severe infection reduces iron absorption very markedly, sometimes to only $\frac{1}{10}$ of the normal. The daily requirement in the adult male diet, i.e., the amount required to replace the small amounts lost in the excreta and in discarded cells from the skin and gastrointestinal tract, is between 5 and 10 mg. For growing children and for women, owing to losses due to menstruation, and especially during the later months of pregnancy, the requirement is double this amount. The iron content of the average daily diet is around 5 mg. The incorporation of iron in the hemoglobin molecule is a function of the reticulocytes, and in cells of the earlier stages of erythropoiesis, iron is not taken up by the erythrocytes.

A conception of the mechanism whereby the iron absorption is graded, as described above, in conformity with the state of the iron stores has been offered by Granick and his associates. Their results with radioactive iron indicate that ferrous iron passes from the intestinal lumen into the mucosal cells, and is here oxidized to the ferric form (FeOH). The ferric iron combines with a protein

known as *apoferritin* (mol wt around 460,000) the iron-phosphorus-protein complex *ferritin* being formed. Ferritin is thus a conjugated protein containing about 23 per cent of iron. The iron-phosphorus portion has the approximate formula $[(\text{FeOOH})(\text{FeOPO}_3\text{H}_2)]$. When apoferritin becomes fully "saturated", iron absorption ceases for the time, the ferritin is in equilibrium with the iron of the plasma. But a fall in the concentration of plasma iron is followed by the separation of the element from ferritin, and its absorption into the circulation, where it combines with β -globulin to form a compound called *siderophilin*. In its separation from ferritin it is converted by reducing substances in the mucosal cell to the ferrous form, but is oxidized in the plasma, and is in the ferric state in siderophilin. The apoferritin remaining after its iron has been removed is free to combine again with iron entering the cell from the intestinal lumen.

Ferritin is also, apparently, the form in which iron is stored in the tissues (chiefly reticuloendothelial cells of bone marrow, spleen and liver). Thus there exist three forms of iron equilibrated with one another—in the intestinal lumen, plasma and the iron depots—which provide for the absorption, transport and storage of the element. Little is known concerning the history of iron from its deposition in the tissues and its incorporation into the hemoglobin molecule, but it seems that it must again be converted to the ferrous state before this can occur. Ferritin has been found to possess vasopressor activity, and has been identified with VDM (see pp 22, 28 and 304). It is also anti-diuretic, though this effect is probably brought about indirectly through the liberation of the hypophyseal diuretic principle.

The *mucosal block theory* of iron absorption just described appears to be protective in nature, preventing the absorption of excessive quantities of the metal, it also helps to account for the "one way" movement of iron mentioned previously. In hemochromatosis (p 79) the mechanism appears to be defective.

Besides its well known function as an essential element in the hemoglobin molecule, and as a constituent of other respiratory pigments, iron appears to play a rôle in the nutrition of epithelial surfaces. Abnormal nail growth, glossitis, fissures around the corners of the mouth and localized thickening of the mucous lining of the esophagus leading to dysphagia occur in anemias due to iron deficiency, and are cured by iron administration.

The erroneous belief has been current in the past that only iron in organic combination could be absorbed and utilized for the construction of hemoglobin. Whipple and associates found that in the posthemorrhagic anemia of dogs iron in any soluble form was utilized, and Elvehjem, Hart and Sherman found that inorganic iron in the form of ferric chloride, pyrophosphate, hypophosphite and glutamate was utilized. The idea that only organic iron is available for hemoglobin synthesis probably arose from the fact that in iron-containing foods (e.g., egg-yolk) which were recognized to be efficient hemoglobin formers, the inorganic iron is masked. It is now known that the truly organic iron of such foods, namely that combined in heme, is not available, since it is not released by peptic or tryptic digestion, whereas the inorganic iron which may represent 50 per cent or more of the total iron of the food is utilizable. This latter after conversion to the ferrous form is estimated by the dipyrindyl method. During digestion the inorganic iron of the food, which is in the ferric state, gives up oxygen to oxidizable substances in the alimentary tract and is converted, in part at any rate, to the ferrous state, as such only is it absorbed. Using radioactive iron as a tracer, Moore and others found that from $1\frac{1}{2}$ to 15 times more of the ferrous form than of the ferric was absorbed from the human gastronomic tract.⁷ Animals fed dipyrindyl become anemic since the ferrous iron is rendered insoluble.

In certain anemias, especially when the gastric juice lacks HCl, which normally aids in the liberation of iron from the food, and facilitates its conversion to the ferrous form, compounds of ferrous iron, e.g., ferrous chloride⁸ and ferrous

⁷ The greater absorbability of ferrous iron as compared with the ferric form is not pronounced in experimental animals.

⁸ Ferric iron, e.g., reduced iron or iron and ammonium citrate, must be given in relatively very large amounts since only a small proportion of it is converted to ferrous iron in the gastro-intestinal tract. When reduced iron is dissolved in dilute hydrochloric acid, ferrous chloride is formed. The administration of a mixture of this sort has been strongly advocated by Lucas and Henderson (Can. Med. Ass. Jour., 1933, 28, 298).

Iron is sometimes given parenterally but it is to be remembered that when free in the blood stream the metal exerts a toxic effect. The lethal dose for an animal is 30 to 60 mg. per kilogram. It has been shown by Castle and associates that in hypochromic anemia a daily dose of 32 mg. of iron in the form of iron and ammonium citrate, administered parenterally, is utilized completely in the formation of hemoglobin. It was also shown that 32 mg. of iron administered in this way is equal to 1000 mg. given by mouth. This is

carbonate are specific in their effects, causing a discharge of reticulocytes from the bone marrow and an increase in the red cell count.

THE SIGNIFICANCE OF COPPER, MANGANESE AND COBALT

Copper is believed to act as a catalyst in some stage of hemoglobin synthesis. It does not itself enter into the structure of the hemoglobin molecule. Waddell, Elvehjem, Steenbock and Hart found that in young rats rendered anemic by being placed upon a diet of whole cow's milk, iron alone failed to promote hemoglobin regeneration.⁹ When the iron was supplemented by a very small quantity of copper, blood regeneration was induced. *Manganese* exerts a similar though less pronounced supplementary effect. The liver is the main storehouse for copper and minute amounts (0.1 to 0.5 mg.) are present normally in blood.

Though anemia due to copper deficiency is unknown in man, a severe and even fatal anemia due to the lack of this metal may occur in farm animals, e.g., the "falling sickness" of South African cattle.

Cobalt has a powerfully stimulating effect upon red cell production and in repeated doses may induce polycythemia (p. 13). This element is also an essential dietary constituent for the synthesis of hemoglobin. Experimental animals kept upon a diet adequate in all respects for normal erythropoiesis, except for the absence of cobalt, develop a severe anemia. Cattle and sheep grazing upon land lacking in this mineral become anemic and fail to respond to iron therapy, but are cured dramatically by the administration of traces of cobalt (see also pp. 13 and 85). Cobalt is a constituent of vitamin B₁₂.

merely a further demonstration of the fact that only a very small proportion of ferric iron is converted into the absorbable form in the alimentary tract and should not be taken in any way to imply that iron administered by injection has any advantage over ferrous compounds given in adequate amounts by mouth. The intravenous administration of iron, it should be remembered, is unphysiological in the sense that the mechanism described on p. 77 is short-circuited. But investigations within recent years indicate that a single dose of about 100 mg. of metallic iron at intervals determined by the severity of the anemia is free from danger of toxic effects or of causing hemochromatosis.

⁹ Some experimenters have obtained a certain degree of hemoglobin regeneration with iron alone, though the rate of regeneration was much increased by the addition of copper.

Hemochromatosis

This is a disturbance in iron metabolism in which extensive deposits of a colloid iron-containing pigment called *hemosiderin* are found in the cells of the liver, spleen and other tissues. Hemosiderin contains up to 55 per cent of iron, but its exact chemical composition is unknown, it probably represents "iron deposited in excess of the capacity of apoferritin to bind" (Drabkin). The total iron content of the body is greatly increased, it may be 10 times the normal amount. A second yellow pigment known as *hemofuscin* is also sometimes present in the connective and muscular tissues. Other features of the condition are *bronzing of the skin*, *cirrhosis of the liver*, *sclerosis of the pancreas* with diabetes (bronzed diabetes), and sometimes testicular atrophy. The disease may appear after repeated transfusions.

Hemochromatosis has been attributed to a high iron content of the diet combined with a deficiency of some essential constituent. A condition resembling hemochro-

matosis has been reported in Negroes of certain South African tribes who subsist mainly upon a diet of maize and ingest excessive amounts of iron derived from the pots in which the food is cooked. It is more probable, however, that the disease in most instances is due to a fundamental fault in iron metabolism and that the mechanism described on p. 77 is in abeyance or defective. A patient suffering from hemochromatosis absorbs 20 per cent of a dose of ferrous iron given orally, or more than 10 times the amount absorbed by a normal person. Yet the quantity found in the blood is less than that present in the blood of a normal person receiving such a dose.

Hemosiderosis is the term applied to the deposit of hemosiderin in the tissues which occurs as a result of the excessive breakdown of red cells in malaria and hemolytic types of anemia. It may be looked upon simply as an exaggeration of the normal process of iron deposition.

CHAPTER 9

THE ANEMIAS

CLASSIFICATION

We have seen that in health the population of red cells and the concentration of hemoglobin in the blood are kept at normal levels by a nice balance between the new formation and the wastage of erythrocytes. Anemia results when the balance is tipped one or the other way, i.e., by a defect of blood formation or an increase in blood wastage. So the anemias may be classified broadly into (A) *those associated with blood loss or increased blood destruction* and (B) *those due to defective blood formation*.

(A) *Anemias due to blood loss or increased blood destruction*

I *Post hemorrhagic anemias*

Hemorrhage

i *Acute*

- ii *Chronic*, as a result, for example of peptic ulcer, uterine bleeding, ankylostomiasis (hook-worm disease), purpura, etc

II *Hemolytic anemias*

Red cell destruction, as a result of

- i *Chemical hemolytic poisons*, lead, arseniureted hydrogen and certain coal tar derivatives
- ii *Specific infections*, e.g., malaria, septicemia
- iii *Abnormal structure of the red cells*, which render them more susceptible to phagocytosis or to disintegration in the blood stream, e.g., chronic congenital hemolytic (Lederer's) anemia, and erythroblastic (Cooley's) anemia
- iv *An endogenous hemolysin of unknown nature*, e.g., chronic congenital hemolytic jaundice and the anemia associated with paroxysmal hemoglobinuria
- v *Anti-Rh, and occasionally anti-A and anti-B agglutinins*, causing hemolytic disease of the newborn (p. 49)

In the hemolytic group the increased blood destruction is manifested by a rise in the concentration of bile pigment in the plasma which gives an indirect van den Bergh reaction (p. 544), a greater excretion of bile pigment and porphyrin in the urine and feces, and the deposit of an iron-containing pigment in the liver and other tissues

(hemosiderosis, p. 79). There is frequently jaundice of a slight or moderate grade.

Sickle-cell anemia is believed to be due to blood destruction as a result of a congenital anomaly of the red cells. In this type, which occurs almost exclusively in Negroes, elongated crescent or sickle shaped birefringent erythrocytes some 15μ or so in length are a characteristic feature. Such cells are found in the blood of a high percentage of Negroes (8 to 9 per cent) though only relatively few, 1 in 40, of these develop anemia. There is also a proportion of Negroes and also a few whites whose red cells become sickle shaped under certain adverse conditions, especially as a result of reduction in the oxygen tension of the blood. This is spoken of as the *sickle-cell trait* and is due to a single sickle cell allelomorphous gene (sickle cell anemia being caused by the presence of two genes). It has been shown by electrophoretic studies that the hemoglobin (probably the globin rather than the heme) of the erythrocytes in sickle cell anemia, in those showing sickle-cells but no anemia, or those in which sickling can be induced, differs somewhat from normal hemoglobin. The abnormal hemoglobin constitutes 100 per cent of the blood pigment in sickle cell anemia, 60 per cent when sickle-cells are present but there is no anemia, and 40 per cent in those in whom sickling can be induced by a low oxygen tension.

Chronic congenital hemolytic jaundice. Synonyms, *acholuric jaundice*, *spherocytic anemia*, *chronic familial jaundice*, *hemolytic splenomegaly*. The chief features of this form of anemia are jaundice and a high incidence of pigment gall stones, *spheroid erythrocytes* (i.e., the diameter of the cells is reduced but their thickness increased, their volume being approximately normal), *reticulocytosis* up to 60 per cent or more of the total red cell population, increased *fragility* of the cells and *enlargement of the spleen*. Splenectomy is usually followed by the disappearance of jaundice, a return of the red cell count to normal and a marked reduction in the reticulocytosis. The spherocytosis gradually disappears and with it usually though not invariably the increased fragility of the erythrocytes. These facts indicate that disordered function of the spleen is not the primary cause of the condition. The cause of the hemolysis is unknown though it may be due simply, as in certain other anemias, to the reaction of the macrophages of the spleen to the defective cells produced by an abnormally functioning bone marrow. Others believe that an endogenous hemolytic agent is responsible, a view which receives support from the experiments of

Dameshek and Schwartz in which a hemolytic serum produced by the injection of guinea-pig's cells into rabbits was employed to produce anemia. The injection of this anti-guinea-pig hemolytic serum into guinea-pigs caused a profound drop in the red cell count and the appearance of cells with spheroid shape, increased fragility and a reticulocytosis. These observers suggest that the greater fragility of the red cells is simply a function of their globular shape and that spherocytosis is a reaction of the bone marrow to excessive destruction of red cells caused by a circulating hemolysin.

Some divide the disease into two distinct forms: congenital and acquired. Justification for the division is provided by the fact that in the *acquired* form the patient's cells, after transfusion into a normal person, have a normal life span, but normal cells transfused into the patient survive only for about $\frac{1}{2}$ the normal time. This points to a hemolysin. In the *congenital* type the behavior of the cells is reversed, the patient's cells injected into a normal subject are removed rapidly but normal cells persist in the patient's circulation for the usual time. This suggests that in the congenital type the abnormality of the cells is due primarily to some defect of bone marrow function. These experiments were performed by Loutit and Mollison, who consider that in the congenital form the spherocytes are unduly susceptible to normal hemolytic processes, that a hemolysin acting under normal conditions may be lysolecithin, and that the slow movement of blood through the spleen favors the hemolytic action. This would explain the benefit which follows removal of the spleen, as well as why this operation does not always reduce the fragility of the cells.

Lederer's anemia is an acute hemolytic anemia. The leucocytes are increased in number and phagocytosis of red cells may be a pronounced feature.

In *Cooley's anemia* the blood contains very thin red cells (*leptocytes*) or elongated forms, and many show concentric rings like a target ("*target cells*") The disease is congenital, being inherited as a Mendelian dominant. There are also splenomegaly and bony changes. It has been suggested that the fundamental abnormality is a defect in the formation of the red cell stroma.

the *B complex*, e.g., pyridoxin, riboflavin, nicotinic acid and folic acid.

iv The macrocytic anemias of sprue and pregnancy are usually classed with the nutritional anemias.

(b) *Lack of or failure in the utilization of the specific anti-anemic factor*—Addisonian pernicious anemia and certain related hyperchromic macrocytic anemias with a megaloblastic type of bone marrow (p. 110). Bone marrow for microscopical examination is obtained by puncturing the sternum, iliac crest or a spinous process with a specially designed styletted needle and aspirating a small sample.

(c) *Macrocytic anemias (hypo- or hyperchromic) with normoblastic or macroblastic bone marrow*

(d) *Toxic agents which induce aplasia of the bone marrow or depress its function—aplastic anemias*. Among such agents are quinine, gold salts, benzol, arsphenamine, radium, X-rays, and sometimes bacterial and syphilitic toxins. The red marrow is greatly reduced in amount, being replaced by fatty tissue. Blood formation is profoundly depressed. Anemia of the aplastic type may also result from exhaustion of the bone marrow following a long period of overactivity induced by some other type of anemia, or may appear without a known cause (*idiopathic aplastic anemia*, p. 90).

The anemias associated with certain conditions, e.g., nephritis and chronic infections, are believed to be due to a relatively mild degree of depression of bone marrow function or of iron absorption. In such types, destruction of red cells by toxic substances (metabolic or bacterial) may also be a contributory factor.

(B) Anemias due to defective blood formation

(a) Nutritional anemias

i *Iron deficiency*—hypochromic, microcytic anemias. The deficiency may be due either to the excessive loss of iron from the body as in chronic hemorrhage or to an inadequate quantity of this element in the diet (p. 76).

ii *Protein deficiency*, though in itself a less common cause of anemia, is not infrequently a contributory factor.

iii *Lack of vitamin C* or of certain factors of

The cause of *splenic anemia* (Banti's) is unknown, it is therefore difficult to fit this type into any of the preceding categories. According to some there is increased blood destruction, but the chief factors appear to be depressed marrow function and the loss of blood resulting from repeated gastro-intestinal hemorrhages. An anemia very closely resembling Addisonian or true pernicious anemia results from infestation with the fish tapeworm, *Diphyllobothrium latum*. A toxin derived from the worm itself is a factor but apparently not the sole one in the development of the anemia. It is exceedingly rare on this continent.

Blood indices The following indices calculated from the hemoglobin concentration, red cell count and cell volume of a specimen of blood are employed to express the characters of the individual cells in the different types of anemia.

The **color index** This is a numerical expression of the hemoglobin content of the individual red cells. It is obtained by dividing the hemoglobin value in grams per 100 cc. by the red cell count, both values being expressed as percentages of the normal (The normal weight of hemoglobin per 100 cc. is taken as 14.5 grams and the red cell count as 5 million.) Thus, if the hemoglobin is 60 per cent of normal and the red cell count 50 per cent, then $60/50 = 1.2$ color index. But if the hemoglobin percentage is reduced to a greater degree than the red cell percentage the index will be less than unity. That is, each red cell contains less than its normal quota of hemoglobin. On the other hand, the hemoglobin concentration of the blood may be greatly reduced but if the reduction runs parallel with the reduction in red cell percentage the index will have the normal value of 1.0.

The **iron index** is obtained by dividing the number of milligrams of iron in 100 cc. of blood by the first three figures of the red cell count. The normal value is between 9.50 and 10.00. Thus if there are 50 mg of iron in 100 cc. of blood and the red cell count is 5,000,000, the iron index is $50/5.00 = 10.00$. Sometimes what is known as the **iron color index** is employed. This is the quotient obtained by dividing the iron content expressed as a percentage of the normal by the percentage of red cells. Thus

$$\frac{\text{Percentage of iron}}{\text{Percentage of red cells}} = \text{iron color index}$$

The **volume index** is an expression of the average size of the red cells in a given sample of blood as compared with the average normal size. It is obtained by dividing the volume of the red cells (as determined by the hematocrit and expressed as a percentage of the volume of a healthy person) by the number of red cells as a percentage of the normal (the normal being taken as 5 million). Thus, if the volume of cells in the sample is 40 cc. per 100 cc. of blood, i.e., 90 per cent of the average normal volume (which is 45 cc.), and the red cell count is 4.5 million, which is also 90 per cent of the normal, then the volume index is $90/90 = 1$. If, on the other hand, the volume is say 70 per cent of the normal and the red cell count 50 per cent then $70/50 = 1.4$ volume index. That is, the average size of the red cells is greater than normal.

The **saturation index** This is an expression of the concentration of hemoglobin in the red cell and is obtained by dividing the figure for the weight of hemoglobin per 100 cc. by the figure for the packed cell volume, both figures being percentages of the normal. Thus, if the weight of hemoglobin is 11.6 grams per 100

cc. (i.e., 80 per cent of normal) and the packed cell volume is 40 (i.e., 88 per cent of normal which is taken as 45) then, $80/88 = 0.90$ saturation index.

HYPCHROMIC MICROCYTIC ANEMIAS—IRON DEFICIENCY

In this group the essential defect is one of hemoglobin formation. The hemoglobin percentage of the blood is reduced to a greater extent than the number of red cells. These, indeed, may show only a slight reduction. The color index is therefore considerably below the normal, which means that each red cell has received less than its normal quota of pigment. The erythrocytes are also smaller than normal, so the volume index is also low. The low color index is in part the result of the smaller size of the red cell but also of a reduced concentration of pigment throughout the red cell's substance. The saturation and iron indices are therefore also lowered. Some of the corpuscles are so pale that they resemble "ghosts" (p. 63) or only the peripheral zone of the cell is colored (anisochromasia) (See figs. 9.1 and 9.2, and frontispiece).

Iron deficiency is considered to be the essential cause of the anemias belonging to this class. The deficiency may result from an inadequate amount of iron in the diet or from defective absorption of the metal from the food.

Chlorosis is a type of hypochromic anemia which until the last 40 years or so was commonly seen in girls and young women between the ages of 16 and 22 years. It received its name from the peculiar greenish pallor of the skin (green sickness). There was usually gastric hyperchlorhydria. The disease is now rare, possibly owing to the more adequate diets of the present generation.

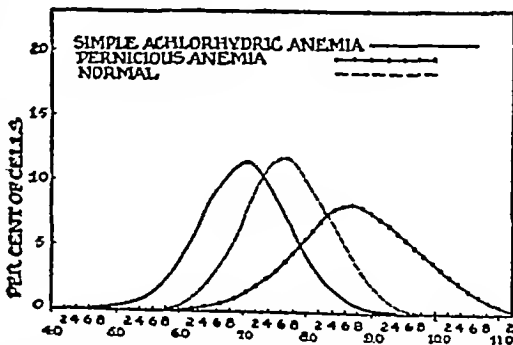


FIG. 9.1 Red-cell-diameter distribution curves in simple achlorhydric anemia and in pernicious anemia compared with the normal (after Haden, see also fig. 9.2).

HYPOCHROMIC ANEMIA WITH HYPOCHLORHYDRIA OR ACHLORHYDRIA (ALSO CALLED IDIOPATHIC HYPOCHROMIC ANEMIA)

Some believe that the failure of gastric function is the primary cause of this type. The hydrochloric acid of the gastric juice is normally an important factor in the liberation of iron from the food and in its conversion to the ferrous state. These are necessary preliminaries for its absorption. According to Davidson and Leitch the hypochlorhydria is merely an accessory factor in the production of the iron deficiency, the iron intake being also below the normal. In the experience of these observers, the diets of patients showing this variety of anemia were poor in first class protein, milk and vegetables and contained about half the minimal daily iron requirement of from 15 to 20 mg. Repeated small hemorrhages or excessive menstrual losses are often contributory factors. It is seldom that one or more of the causes enumerated cannot be detected and for this reason the title "idiopathic" as applied to this type of anemia is scarcely appropriate. Hypochromic anemia sometimes follows operation for the removal of a large section of the stomach.

HYPOCHROMIC ANEMIA WITHOUT HYPOCHLORHYDRIA

Hypochromic anemia with normal gastric acidity also occurs and like the preceding is due to iron deficiency. It occurs most frequently in women of childbearing age when the dietary deficiency is aggravated by the losses of iron incident to menstruation or repeated pregnancies. Davidson has recently emphasized the need for larger amounts of iron during this period in women.

ANEMIA OF INFANTS

The fetus accumulates a store of iron in the liver in the later months of gestation which serves as a reserve which is drawn upon for the manufacture of hemoglobin in infancy. The high red cell concentration (p 11) with which the infant comes into the world also contributes to the iron reserves. In the normal infant the iron stores are sufficient for the manufacture of hemoglobin for the first six months or so. Growth, however, makes heavy demands upon the iron supplies and after the first half year it is necessary to provide a diet which will contain adequate amounts of iron in order to guard against the development of anemia of the hypochromic type. Milk, it will be recalled, is very poor in both iron and copper. The develop-

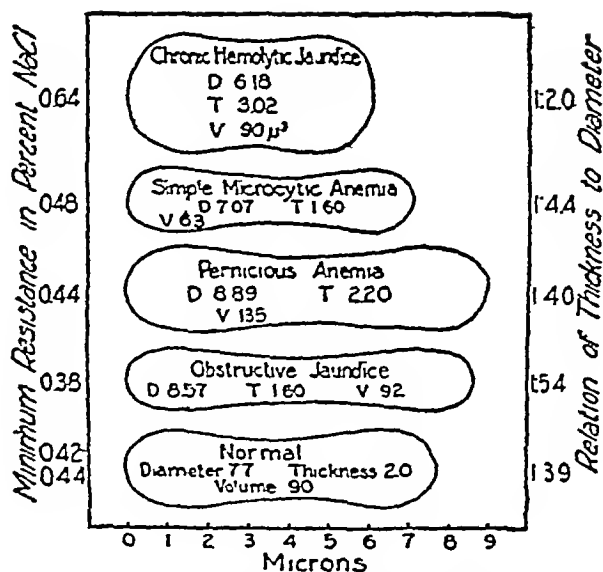


FIG 92 Red cell diameter, thickness and volume in various clinical conditions compared with the normal (after Haden). D, diameter, T, thickness, V, volume

ment of anemia in milk-fed rats has been mentioned and the anemia of sucklings is a problem in the breeding of farm animals. If the iron stores are deficient at birth as in premature infants or as a result of maternal anemia, anemia may occur in the very young infant.

The hypochromic anemias respond in a spectacular fashion to the administration of inorganic iron (e.g., ferrous chloride, ferrous carbonate, etc., see p 78). The administration of copper is rarely necessary since this occurs in sufficient quantity in the diet and as an impurity in iron preparations.

A macrocytic anemia with a megaloblastic type of bone marrow also occurs in infants. It responds to liver extract (p 84) and to folic acid (p 88). It is frequently accompanied by leukopenia and splenomegaly.

PERNICIOUS ANEMIA (ADDISON'S ANEMIA) AND RELATED MACROCYTIC ANEMIAS

Pernicious anemia is due essentially to a defect in the formation of the red cells, hemoglobin synthesis is unaffected.

The CHIEF FEATURES of the blood picture are

(1) *Great reduction* in the number of red cells and consequently in the hemoglobin percentage. The blood count in a very severe case may be less than 10 per cent of the normal.

(2) The red cells are reduced in number below the normal to a greater extent than is the hemoglobin percentage. The *color* and *volume indices* are therefore *raised* above normal, but the *saturation* and *iron indices* are

normal, the high color index being due to the greater size of the cells

(3) Large cells—the average diameter of the cells is increased to between 8 and 9 microns and exceptionally large cells called *macrocytes* are plentiful. The average volume of the individual red cells (figs 9.1 and 9.2) is about 135 cu. microns (normal about 90 cu. microns). *Normoblasts* and earlier forms are present in the circulation, and a characteristic large nucleated cell containing basophilic material in its cytoplasm and little or no hemoglobin is a feature of the blood picture. This cell, known as the *megaloblast*, differs chemically and morphologically from any normal cell of the erythrocyte series found either in bone marrow or blood. The reticulocytes are around 3 per cent.

(4) The total number of leukocytes is reduced but the lymphocytes are relatively increased

(5) Great variation in the size of the cells—*anisocytosis*, the cells varying from those smaller than normal to the large cells mentioned above. *Poikilocytes* are moderate in number

(6) Increase in iron and bilirubin of the plasma, increased excretion of pigment (urobilin)—indirect van den Bergh

(7) Blood volume reduced mainly as a result of the red cell diminution, the plasma volume being around the normal level

(8) Fragility of the red cells usually slightly reduced.

OTHER FEATURES

(1) The red bone marrow is hyperplastic. It extends into the shafts of the bones displacing the yellow marrow and even the bony walls may be eroded. Upon microscopical examination the marrow shows megaloblasts and other immature forms in large numbers (see p 109 and Plate I)

(2) Achlorhydria almost always exists

(3) Sore, shiny tongue, *glossitis*, atrophy of lingual papillae.

(4) Chronic combined degeneration of the cord

(5) Urobilin appears in the urine in severe cases, and in all those in which plasma bilirubin is increased urobilinogen is in excess in the feces. In health from 1/140 to 1/340 of the total amount of hemoglobin in the blood is excreted daily as urobilinogen. In pernicious anemia 1/10 of the total hemoglobin may be excreted as urobilinogen. Hemosiderin deposits (p 79) occur

(6) The disease shows remissions and relapses. During the remissions, the blood picture approaches the normal and the percentage of reticulocytes increases—*reticulocyte crises*. During the relapses, the characteristic hematological features of the condition are exaggerated.

The essential factor in the production of pernicious anemia is not believed to be increased blood destruction, but rather a defect in blood forma-

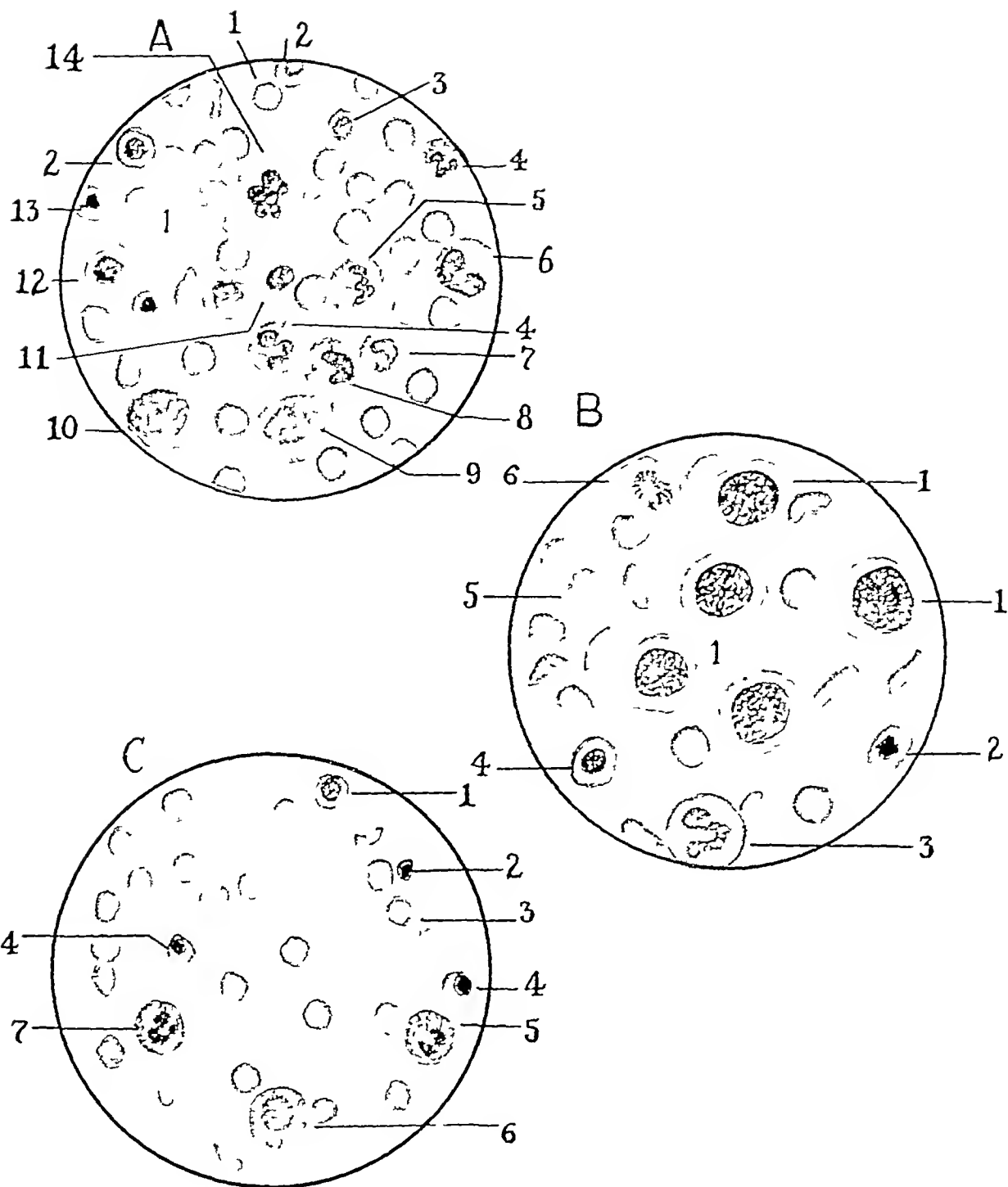
tion—the reversion of the bone marrow to a more primitive type of erythropoiesis (see p 109). Increased blood destruction undoubtedly occurs as evidenced by the rise in plasma bilirubin, but it is a secondary effect. The abnormal erythrocytes probably stimulate the phagocytic activities of the reticulo-endothelial cells (p 104) in the spleen, liver and elsewhere. Since the fragility of the red cells is reduced rather than increased it would not appear that they disintegrate more readily in the blood stream (see p 73). The discovery of an anti-anemic principle in liver and subsequent researches arising from the employment of this principle in the treatment of pernicious anemia has shed a flood of light upon the nature of the disease.

TREATMENT WITH LIVER, LIVER EXTRACT AND GASTRIC TISSUE

In 1926 Minot and Murphy, inspired by the laboratory finding of Whipple and his associates, namely, that liver was the most effective article of diet for the treatment of anemia in dogs, tried the effect of adding liver to the diet of pernicious anemia patients. The spectacular success which followed this treatment is well known and today lightly cooked liver (from $\frac{1}{2}$ to 1 pound per day) but preferably an extract of liver, or the essential principle, vitamin B₁₂, is recognized as a specific for the disease. Kidney tissue was shown to have a similar though less pronounced curative effect. The hematopoietic substance present in the liver and in non-protein extracts of liver tissue and which is effective in the treatment of pernicious anemia is known as the *anthanemic* or *hematinic* principle, or perhaps more generally today as the *erythrocyte maturation factor* (EMF), or, again, as B₁₂ (see below). The action of the hematinic principle is several times more potent given parenterally than when given orally. It may be mentioned here that Wilkinson found that the anti-anemic principle was present in normal human liver and in the livers of pernicious anemia patients who had received specific treatment, but was absent from the livers of untreated subjects of the disease.

The chemical identification of the erythrocyte maturation factor of liver, B₁₂

Nearly 30 years have passed since the discovery of an anti-anemic factor in liver and, though investigations since then have been prosecuted assiduously in several laboratories, only recently has a near approach to its exact chem-



Samples of bone marrow obtained by sternal puncture

A Normal

- 1 Erythrocytes
- 2 Erythroblasts
- 3 Early normoblast
- 4 Neutrophil leucocyte
- 5 Eosinophil leucocyte
- 6 Neutrophil metamyelocyte
- 7 Basophil metamyelocyte
- 8 Eosinophil metamyelocyte
- 9 Neutrophil myelocyte
- 10 Myeloblast
- 11 Hemocytoblast
- 12 Lymphocyte
- 13 Late normoblast
- 14 Megakaryocyte

B In pernicious anemia

- 1 Megaloblasts
- 2 Late normoblast
- 3 Giant neutrophil leucocyte
- 4 Erythroblast
- 5 Erythrocytes (macrocytes)
- 6 Megaloblast in mitosis

C In microcytic anemia due to iron deficiency

- 1 Early normoblast
- 2 Erythrocytes (microcytes)
- 3 Extruded nucleus from normoblast
- 4 Late normoblasts
- 5 Eosinophil leucocyte
- 6 Myelocyte
- 7 Neutrophil leucocyte

ical structure been made. Early in 1948 Rickes and his associates reported the isolation of red, needle-shaped crystals of a substance which was many times more potent than the most purified extracts of liver, causing a reticulocyte response in pernicious anemia patients in a parenteral dose of 1 microgram, or even less. Progress in the concentration of this highly potent material was guided by the observation made previously that a substance which stimulated the growth of *Lactobacillus lactis* Dorner was closely associated with the hematopoietic factor of liver. In the same month Smith reported the isolation from proteolyzed liver of an amorphous red principle which was effective in pernicious anemia in doses of around half a milligram. Smith described his red product as "having about the color of cobalt salts." The red crystalline material, now generally referred to as vitamin B₁₂, has since been shown to be a cobalt complex, and is apparently the erythrocyte maturation factor in pure or nearly pure form. It also contains phosphorus and nitrogen, but no sulfur, its molecular weight is around 1600 and its approximate formula is C₆₁₋₆₄H₈₆₋₉₂N₁₄O₁₃PCo. A cyano group is coordinately bound to the cobalt atom. Its chemical designation is therefore cyanocobalamin.

Several analogues of this compound have been discovered and called vitamins B_{12a}, B_{12b}, B_{12c}, and B_{12d}. B_{12a} was originally prepared by a reaction of B₁₂ and hydrogen over a platinum catalyst, and later from cultures of *Streptomyces griseus*. B_{12b} was obtained from cultures of *Streptomyces aureofaciens* but it has since been found that B_{12a} and B_{12b} are identical substances, in which the cyano group of B₁₂ is replaced by an hydroxo group. Their chemical name is *hydroxocobalamin*. B_{12c} is also obtained from cultures of *Streptomyces griseus*, it and B_{12d} are closely similar, but are distinguishable by their absorption spectra. All of these compounds have the same action as vitamin B₁₂ itself. Vitamin B₁₂ is produced in the human colon, even of pernicious anemia patients, by bacterial action, but it is not absorbed. It is also found in activated sewage sludge, in cow dung, beef muscle and several other foods, and in the juice of autoclaved fish. The vitamin from some of these sources is conjugated with peptides, and is therapeutically inert but can be activated by digestion with pancreatic juice, or by treatment with potassium cyanide which apparently frees the vitamin from the peptides (see Haumann and Mulli). A process has been developed for the production of vitamin B₁₂ through fermentation by *B. megatherium*.

It may be mentioned that the cobalt ion by itself elicits no hemopoietic response in pernicious anemia.

Besides its importance in hemopoiesis and its action in promoting the growth of certain bacteria, vitamin B₁₂ is a growth factor for higher animals, including man. It also has an influence upon the metabolism of certain sulfur containing amino-acids (e.g., conversion of homocystine to methionine). According to Niewig, vitamin B₁₂ is essential for the synthesis of ribonucleic acid (ch 48), a constituent of the nuclei of all cells, this would account for the neurological defects in pernicious anemia, and for the abnormalities of certain epithelial surfaces, such as that of the tongue, as well as for the failure in erythrocyte maturation.

Vitamin B₁₂ is by far the most potent hemopoietic principle known, being several thousand times more potent weight for weight than folic acid. B₁₂ is also effective in arresting the course of the neurological lesions of pernicious anemia.

The production of the anti-anemic principle, gastric (intrinsic) and extrinsic factors

As already mentioned, achlorhydria is virtually an invariable accompaniment of pernicious anemia. The significance of this fact was demonstrated in 1929 by Castle who found that the gastric contents of a normal person during the digestion of meat were curative when fed to a subject of pernicious anemia. Later Castle and his associates showed that pure gastric juice obtained from a normal person by the administration of histamine when incubated with beef-steak produced the curative material. The active principle was not produced when beef was incubated with gastric juice of a patient with pernicious anemia. The production of the anti-anemic principle is not due to the action of hydrochloric acid, pepsin, rennin or lipase but to the presence of an enzyme-like but unidentified substance which acts at a pH of 7 (the optimal pH for the action of pepsin is 1.6). Before this, Sturgis and Isaacs found that gastric tissue contained the material necessary for the formation of the anti-anemic factor, it being, like liver itself, effective in the treatment of pernicious anemia. Desiccated de-fatted hog stomach is, therefore, also employed as an alternative to liver or liver extract for oral administration. The gastric factor is less stable than the liver principle, being destroyed by temperatures above 45°C or by digestion with pepsin or trypsin. In the hog, the gastric factor is produced by the mucosa of the pyloric and cardiac regions of the stomach and the commencement of the duodenum, i.e., regions which

secrete an alkaline juice (pyloric, cardiac and Brunner's glands) In the human subject the factor, according to Fox and Castle, is formed in the fundus of the stomach, none is found in the pyloric region nor in the duodenal secretions

Two factors, therefore, were recognized as being concerned in the hematopoietic action of liver extract, one in the food, especially in such items of diet as are rich in the vitamin B complex, it was called the *extrinsic* factor, the other factor is secreted in the gastric juice and is known as the *intrinsic*¹ factor

The theory which, until recently, seemed best to fit these facts and to relate them to the specific hematopoietic action of liver extract is the following² In health the essential anti-anemic principle is derived from food through the action of the "intrinsic" (gastric) factor upon an "extrinsic" factor contained in the diet. The anti anemic principle so formed was then absorbed and stored in the liver and possibly other organs, to be drawn upon for the maintenance of normal activity of the erythropoietic tissue (bone-marrow) Pernicious anemia followed if the essential enzyme were absent from the gastric juice The anti-anemic principle could, therefore, be administered preformed, as in liver extract, or gastric tissue together with a good diet could be given from which the patient manufactured his own hematonic principle

A characteristic lesion is found in pernicious anemia which readily explains the achlorhydria, and the lack of the intrinsic factor The fundus and body of the stomach show atrophy of the mucosa and extreme thinning of all coats The gastric glands are almost completely destroyed, the muscular coat is atrophic The pyloric region, which does not produce the intrinsic factor, is normal One would be led to expect that total gastrectomy in man would be followed by pernicious anemia, yet this is not the case Only seldom has such a result been reported The hypochromic type, or a macrocytic anemia without a megaloblastic marrow, is more frequently a sequel of gastrectomy Even removal of the entire stomach from dogs is not followed by pernicious anemia. The reason that pernicious

anemia does not follow removal of the entire stomach has received no entirely satisfactory explanation The most probable reason is that the survival time (1-2 years) following the operation in man (usually for carcinoma) is too short for the development of the disease

The extrinsic factor identified as B₁₂, the action of the intrinsic factor

The theory given above has, in the main, been substantiated by later work, but certain important emendations are required, especially as to the action of the intrinsic factor in gastric juice The extrinsic factor for many years eluded identification Though known to be present in beef, liver, rice polishing and the vitamin B preparation known as marmite (autolyzed yeast), more precise knowledge of its nature could not be secured Castle believed it to be a factor of the B complex, or a closely associated substance This has proved to be so, but most surprisingly it turns out to be the hematonic principle (B₁₂) itself, a discovery which demands the abandonment of the previous theory in so far as it relates to the supposed mechanism of production of the hematonic principle The long search for the extrinsic factor then has ended But why does pernicious anemia ever develop, since any good diet contains more than enough B₁₂ for normal hematopoiesis It is also present (probably as a result of bacterial action) in the contents of the large intestine of normal persons, as well as of pernicious anemia patients An extract of feces of a subject of pernicious anemia causes a hematopoietic response when injected into another patient with the disease The intrinsic factor, it is now revealed, is essential for the adequate *absorption* of vitamin B₁₂, exerting most probably a catalytic action upon the absorptive process In the absence of the intrinsic factor there is "starvation in the midst of plenty" For this reason B₁₂ is effective given parenterally in pernicious anemia in a fraction of the dose that is required by mouth However, if B₁₂ is administered with normal gastric juice the size of the effective oral dose is much reduced In concluding this section the following remarks of Castle may be quoted

"The disease would not develop if the patient could effect daily the transfer of a millionth of a gram of vitamin B₁₂ the distance of a small portion of a millimetre across the intestinal mucosa and into the blood stream Thus he cannot do, principally as a result of failure of his stomach to secrete into its lumen some essential but still unknown substance Yet the patient may each day absorb without much difficulty the prod-

¹ The intrinsic factor is also known as *hemopoietin* (Wilkinson)

² This theory implies, of course, that the gastric defect is the primary cause of the blood condition Yet, though achlorhydria is present in about 5 per cent of persons, only a small proportion develop pernicious anemia, indeed the hypochromic (p 82) type of anemia is more likely to result The probable reason is that in the cases of achlorhydria in which pernicious anemia does not occur, the intrinsic factor is not lacking Some observations of Castle, who found the intrinsic factor present in cases of achlorhydria without anemia, are in accord with this explanation

ucts of the digestion of many grams of carbohydrate, fat and protein from foods that in addition may contain consequential amounts of vitamin B₁₂ in terms of his trivial need "

Other megaloblastic anemias

Though pernicious anemia is the most common type of anemia showing a megaloblastic bone-marrow, there are other forms with closely similar marrow and blood pictures but whose pathogeneses are not altogether clear. Such megaloblastic anemias occur in *tropical sprue*, *pregnancy*, *carcinoma of the stomach*, *gastrocolic fistula* which short circuits the region of the small intestine from which B₁₂ is absorbed, and *infestation with the tapeworm *Diphyllobothrium latum**. The anemia of sprue, in which the absorption of fat is defective, is due possibly to the interference with the absorption of vitamin B₁₂ other than through a lack of the intrinsic factor, though in certain cases the latter is absent even though there is not achlorhydria. In the megaloblastic anemia of pregnancy the marrow and blood pictures are similar to those of pernicious anemia, but there is as a rule not true achlorhydria, nor degenerative changes in the spinal cord, recovery follows childbirth. Also, many cases of this disease, like that described by Wills as occurring in the tropics, and called *tropical nutritional (or macrocytic) anemia*, are unresponsive to the administration of vitamin B₁₂, but respond readily to crude liver extracts or to folic acid.³ In carcinoma of the stomach there is achlorhydria and the anemia may be indistinguishable in every way from pernicious anemia, even to the neurological symptoms, it is probably produced in the same manner—lack of the intrinsic factor. In the anemia caused by *Diphyllobothrium latum*, the intrinsic factor is produced as usual, but its activity is inhibited presumably by a principle produced by the worm. The investigations of von Bonsdorff indicate that the parasite must be situated high up in the small intestine in order to cause anemia. Finally, a rare type of megaloblastic anemia which fails to respond to the hematinic principle was described originally by Wilkinson and Israels. It is not due to the lack of the erythrocyte maturation factor (B₁₂), but to failure of the latter to be utilized by the bone marrow. It was, therefore,

named *achrestic anemia* (Gk, *a*, privative + *chresis*, a using)

The response to liver or gastric tissue

The anti-anemic principle acts upon the bone marrow, restoring the blood-forming processes to normal. The maturation of the erythrocytes is hastened, the primitive cells of the marrow—megaloblasts—disappearing to be replaced by cells of later stages of development, e.g., late erythroblasts and normoblasts (see p. 109). This specific effect of the active principle was demonstrated some years ago by Sabin who found that the addition of liver extract to the early chick embryo accelerated the division and maturation of the primitive megaloblastic cells of the blood islands. The direct effect of B₁₂ upon erythropoiesis has been demonstrated more recently in man, when the vitamin is injected into megaloblastic bone marrow, the abnormal erythropoiesis is corrected locally before a general effect occurs (i.e., on the marrow in other situations).

The first detectable effect of specific treatment in pernicious anemia is a rise in the reticulocytes. In untreated cases of pernicious anemia these constitute 2 per cent or less of the red cells. Within from 2 to 5 days, after vitamin B₁₂, a potent liver extract, or a preparation of gastric tissue has been administered, large numbers of reticulocytes appear. The increase reaches its maximum about the fifth day, when the percentage is from 10 to 40 per cent. From then on the reticulocyte population declines, the immature cells having undergone maturation in the blood stream (fig. 9.3). This is reflected in an increased number of mature erythrocytes. The lower the red cell count before treatment, the greater is the reticulocyte response to specific therapy. This fact is interpreted as indicating that in the more severe forms of the disease, megaloblasts constitute a large percentage of the marrow population, the antianemic principle hastens the development of these primitive forms to reticulocytes which are then promptly expelled into the circulation. When, on the other hand, the anemia is less severe the percentage of later forms, e.g., normoblasts in the marrow is greater and the stimulus to maturation exerted by the liver principle carries them to the adult stage. A larger proportion of mature erythrocytes is therefore discharged into the circulation. In cases in which the red cell count is 3 million or over only a slight rise in the reticulocytes occurs, the rise in the red cell

³ Tropical nutritional anemia occurs most frequently and severely in pregnant women. An anemia with similar characters, and which is resistant to treatment with purified preparations of liver, but is amenable to crude liver extracts or folic acid can be produced in monkeys.

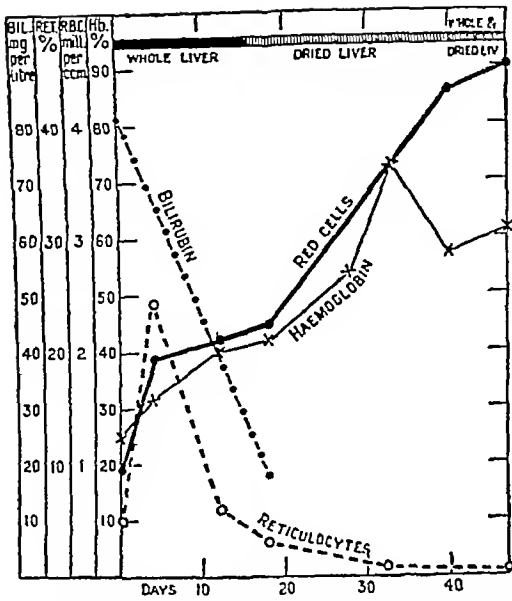


FIG 9-3 Showing effect of specific liver therapy upon the reticulocytes, erythrocytes, hemoglobin and plasma bilirubin (after Dyle)

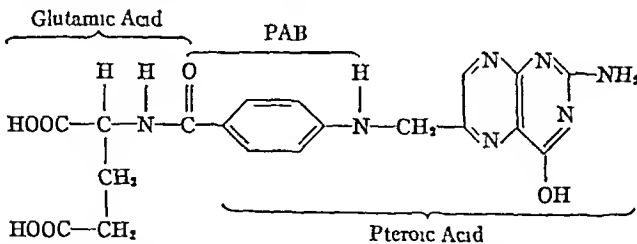
count being then due almost entirely to an increase in the number of mature cells

The rise in the hemoglobin concentration of the blood lags behind the multiplication of the red cells, so the color index returns to normal within a short time. If the stimulus to erythropoiesis is intense, the hemoglobin cannot be manufactured in sufficient quantities to furnish each cell with its

blood picture. The neurological symptoms may improve if vitamin B₁₂ is given early, and in a dosage considerably greater than that required for maintaining the normal blood picture (ch 66). The degenerative changes in the cord are due to a lack of the hematinic principle, and are not simply secondary to and caused by the anemia.

Pteroylglutamic acid Synonyms: *folic acid*, vitamin B₉⁴, liver *Lactobacillus casei* factor, norite eluate factor, vitamin M, Wills factor, factor U

In 1941 Mitchell and his associates obtained a substance with acidic properties from the green leaves of various plants, e.g., spinach, peas, clover, etc., and gave it the appropriate name of *folic acid*. What has since proved to be the same principle had previously, from microbiological studies, been found in concentrates of liver and yeast (*norite eluate factor*). It stimulates the growth of certain microorganisms, e.g., *Lactobacillus casei*, *Streptococcus faecalis* R (or *S. lactis* R), and is essential for the existence of certain protozoa and insects. It is, therefore, also known as the *liver L. casei factor*, but is now more commonly referred to by its chemical name *pteroylglutamic acid* (PGA). It is made up of a pteridyl ring joined through para-aminobenzoic acid to a molecule of glutamic acid. Para-aminobenzoic (PAB) acid appears to be a precursor from which pteroylglutamic acid is synthesized in nature. The chemical structure of pteroylglutamic acid is shown below.



Pteroylglutamic acid, N-[4-[(2-amino-4-hydroxy-6-pteridyl)methyl]amino]-benzoyl] glutamic acid

quota of pigment and the color index falls well below the normal value.

With the improvement in the blood picture the general symptoms of the disease abate, but the secretion of acid gastric juice is rarely if ever restored. Specific treatment therefore does not remove the primary cause of the disease, but must be persisted in for the rest of the patient's life. The maintenance dose is, of course, much less than that required originally for the restoration of the normal

Of fundamental importance in the physiology of pteroylglutamic acid is its indispensability for the synthesis of desoxyribonucleic acid, a constituent of the nuclei of all cells (ch 48). Certain analogues

⁴The subscript "c" was adopted to designate a crystalline B factor in liver, since identified as folic acid, which prevented a macrocytic anemia and leukopenia in the chick. Vitamin M referred to a crystalline factor, also isolated from liver, which prevented anemia, leukopenia and a sprue like condition in monkeys.

of folic acid, such as, *aminopterin* (4-aminopteroylglutamic acid, or 4-aminofolic acid), *amethopterin* (4-amino-10-methyl pteroylglutamic acid) *amino-au-fol* (4-aminopteroylaspartic acid) antagonize this action of the vitamin. Pteroylglutamic acid is synthesized in the intestine of some mammals, especially of the rat, and the analogues are useful in producing folic acid deficiency for experimental purposes. They have also been used in the treatment of acute leukemia, especially in children: the number of leukemic cells is reduced, and temporary remissions of the disease induced. A permanent cure does not, however, result, and the administration of the compounds is attended by certain toxic symptoms, e.g., stomatitis, diarrhea, alopecia and deafness.

The hemopoietic action of pteroylglutamic acid

Several observations had pointed to the existence in liver, or in crude liver preparations, of a principle effective in certain macrocytic anemias which failed to respond to highly purified liver extracts. There also appeared reports of a beneficial action of preparations of brewer's yeast, which does not contain vitamin B₁₂, in macrocytic anemias and in the nutritional anemia of monkeys (p. 87). Suspecting that the unknown hemopoietic substance in crude preparations of liver was "folic acid", Spies and his associates administered it to pernicious anemia patients with remarkable success. Pteroylglutamic acid is also effective in other anemias characterized by a megaloblastic bone marrow, e.g., the anemias of sprue and pregnancy, but fails in the treatment of macrocytic and other types of anemia with a normoblastic or macroblastic bone marrow. It relieves the nutritional leukopenia of monkeys and stimulates the production of granulocytes, and is therefore of value in the treatment of agranulocytosis (p. 101). *It has no effect whatever in arresting the progress of the neurological lesions (chronic combined degeneration of the cord) in pernicious anemia.* Indeed, the neurological signs appear to be aggravated by its administration. Its action in restoring the bone marrow to normal is identical with that of the vitamin B₁₂.

Pteroylglutamic acid was synthesized by Angier and associates in 1945. The synthetic compound is as therapeutically effective as the natural principle.

The chief sources of "folic acid" are liver, kidney and green vegetables, its concentration in plant foods appears to run parallel with their chlorophyll content.

Conjugates of pteroylglutamic acid A compound having three molecules of glutamic acid instead of one is produced during fermentation by a diphtheroid organism (*Corynebacterium*). This conjugate of "folic acid" is known as the *fermentation Lactobacillus casei* factor, or chemically as *pteroyl-triglutamic acid* (or pteroyldiglutamyl glutamic acid). A second conjugate, isolated in crystalline form from liver and yeast by Pfiffner and his associates, has seven glutamic acid molecules, i.e., six more than "folic acid", and is called *B₆ conjugate*, or *pteroylheptaglutamic acid* (or pteroylhexaglutamyl glutamic acid). The "folic acid" in food is present largely in the form of conjugates. Though normally these are the body's chief sources of pteroylglutamic acid, they are ineffective, uncertain, or have only a slight therapeutic action in pernicious anemia (see below).

The breakdown of the conjugates and the liberation of pteroylglutamic acid in the healthy body is brought about by the action of enzymes known as *conjugases*, present in mammalian liver and kidney, in the pancreas of chicks and in some vegetables, e.g., potatoes.

The possible relationship of the erythrocyte maturation factor of liver to pteroylglutamic acid

The chemistry and actions of "folic acid" and of the anti-anemic factor of liver show that they are separate and distinct principles. Moreover, it soon became evident that the concentration of "folic acid" in highly purified and potent liver extracts is equivalent to only a small fraction of the amount required to elicit a reticulocyte response in pernicious anemia. Again, as mentioned above, pteroylglutamic acid is wholly ineffective in arresting the cord changes of pernicious anemia. The non-identity of the two principles was finally settled by the isolation of crystalline B₁₂. Yet it is hard to believe that a physiological relationship of some sort does not exist between the two principles. The most likely explanation of the functional relationship between folic acid and vitamin B₁₂ is the suggestion that vitamin B₁₂ is necessary in some way for the liberation of pteroylglutamic acid from its conjugates. This might be effected by suppressing the action of conjugase inhibitors, or by aiding in a more direct way the action of these enzymes. Such a mechanism would explain why the patient with pernicious anemia does not receive sufficient pteroylglutamic acid from a diet which is an ade-

quate source of this principle for normal persons. That is to say, had the pernicious anemia patient received and utilized all the "folic acid" present in and derivable from conjugates in an ordinary diet, he would not have developed the disease. On the other hand, a diet obviously deficient in "folic acid" does not cause pernicious anemia, and the quantity of the vitamin required therapeutically is usually in excess of that found *free* in the usual diet. All of which points to some defect in the utilization of *pteroylglutamic acid conjugates*.

Since pteroylglutamic acid exerts no beneficial influence upon the development of cord changes, the view is supported that an action separate and distinct from its hemopoietic effect is exerted upon nervous tissue by B₁₂.

More direct evidence can be cited in support of the conception that "folic acid" and vitamin B₁₂ are two parts of one hemopoietic mechanism. When, for example, B₆ conjugate is fed to a normal person, increased amounts of pteroylglutamic acid appear in the urine, but no such increase is observed in a patient with pernicious anemia and there is no therapeutic effect, though a response followed the administration of "folic acid" itself. When, however, the administration of the conjugate was preceded by treatment with liver extract containing only traces of "folic acid", it was followed by an increase in pteroylglutamic acid excretion almost equal to that which occurs in normal persons. Even more striking results were obtained in the macrocytic anemia of sprue. It was found, furthermore, that, if a pernicious anemia patient is under the influence of the liver principle and given conjugase inhibitor together with B₆ conjugate, he excretes more pteroylglutamic acid than does a patient in relapse.

"*Citrovorum factor*", "*folic acid*", *leucovorin*. A derivative of pteroylglutamic acid which promoted the growth of the micro-organism, *Leuconostoc citrovorum*, was isolated in 1948 by Sauberlich and Baumann from liver extracts and yeast. This substance which is now known as the "*citrovorum factor*" has an action similar to that of pteroylglutamic acid in inducing maturation of megakaryoblasts in pernicious anemia and other megaloblastic anemias, it also annuls the toxic effects of such folic acid antagonists as aminopterin and amethopterin. Substances with the same action, and apparently identical with the "*citrovorum factor*" have been prepared artificially, and named "*folic acid*" (Sive and associates), and *leucovorin* (Jukes and associates).

Other anti anemic agents. The pyrimidine base *thymine* (5-methyluracil) and *xanthopterin* (uropterin)

give a hematopoietic response, but one which is much inferior to that of pteroylglutamic acid. An amount of thymine several thousand times the weight of an effective dose of pteroylglutamic acid is required to produce a reticulocyte response in pernicious anemia (Spies and associates). Xanthopterin is a pigment present in butterfly wings. It is effective in monkeys suffering from nutritional anemia and leukopenia. It also causes a reticulocyte response in the anemia, accompanied by symptoms resembling sprue ("rat sprue") which develops in rats fed exclusively upon goats' milk. Xanthopterin has no demonstrable clinical value.

Macrocytic (hypochromic or hyperchromic) anemias with normoblastic bone marrow

Macrocytic anemias showing a resemblance to Addisonian anemia, in so far as the blood picture is concerned, occur in various diseases, e.g., carcinoma (other than gastric), syphilis, etc. But the bone marrow is of the normoblastic type. Since they are not due to deficiency of the antianemic principle, they do not respond to the administration of liver extract or of gastric tissue. Nor do they respond to the administration of "folic acid".

IDIOPATHIC APLASTIC ANEMIA

This is a comparatively rare type of anemia in which there is a rapidly progressive reduction in all the blood cells—erythrocytes, leukocytes and platelets. There is little or no evidence of blood regeneration, reticulocytes are very scarce and nucleated forms are usually absent. The red cell count may reach an extraordinarily low figure—213,000 per cubic millimeter in a case reported by Ehrlich. Granulocytes and platelets may entirely disappear. The marrow is hypoplastic or aplastic, there is a great reduction in its cellular elements and almost complete absence of hemopoietic activity. The lymphocytopenia which also occurs, but is less marked than the reduction in granulocytes, suggests that the entire hemopoietic system is affected. The causes of the bone marrow hypoplasia have been mentioned (p. 81).

A number of cases have been reported recently (Thompson, Richter and Edsall, Anderson) in which the typical blood picture of aplastic anemia was associated with a normal or even a hyperplastic marrow. In these, to which the term "pseudo-aplastic anemia" might be applied, there would appear to be some interference with the maturation and delivery of the cells into the blood stream rather than to absolute suppression of marrow function.

It is only upon pernicious anemia or macrocytic anemias of the pernicious anemia type with a megaloblastic bone marrow that vitamin B₁₂ or

“folic acid” therapy has any specific effect. Other macrocytic types (with a normoblastic type of marrow), the microcytic anemias and aplastic anemia fail to respond to the administration of liver extract, gastric tissue or pteroylglutamic acid. It has been mentioned that certain megaloblastic

anemias (e.g., that of pregnancy) are resistant to the action of vitamin B₁₂ but respond to folic acid. In post-hemorrhagic anemia and certain other secondary anemias *whole liver* is of value, not from any specific action but simply because it furnishes iron and protein of high quality.

CHAPTER 10

THE WHITE BLOOD CORPUSCLES OR LEUKOCYTES—THE PLATELETS

CLASSIFICATION AND MORPHOLOGY

The white blood cell differs from the erythrocyte in that it contains no hemoglobin, and has a nucleus. The majority of the white cells are also considerably larger than the erythrocytes, measuring from 8 to 15 microns in diameter, the size depending upon the particular variety. They are much less numerous than the red-cells, in the adult they number from 5 to 9 thousand per cubic millimeter of blood. In infancy they are twice as numerous and throughout childhood the count is higher than in the adult. When a film of adult blood is examined under the microscope the white cells appear very sparsely scattered here and there among the crowds of colored corpuscles which outnumber them more than 600 to 1.

On a basis of morphological differences the colorless corpuscles are divided first into two main groups. (I) *Cells with a single nucleus and a clear nongranular cytoplasm*—the lymphocytes and the monocytes, (II) *cells having a lobed or incompletely partitioned nucleus, and a cytoplasm containing fine chromophil granules*—the *granulocytes*. Each of these two main classes are divided further into subgroups on a basis of differences in structure or staining properties¹ (see frontispiece).

I. THE NON-GRANULAR LEUKOCYTES— AGRANULOCYTES

These are of three varieties. (1) *Small lymphocyte*, (2) *large lymphocyte*, (3) *monocyte*. Though these forms show no granules in the protoplasm under the ordinary methods of staining, granulation may be demonstrated after staining with azure-blue. The lymphocytes contain a few coarse azurophil granules, those of the monocytes are fine and very numerous.

(1) **THE SMALL LYMPHOCYTES** These are slightly larger than the red cells—about 8 microns in diameter. The nucleus is relatively large, slightly indented and stains more deeply with basic dyes than the surrounding narrow rim of cyto-

plasm which separates it from the boundary of the cell. The small lymphocytes originate in lymphoid tissue and are found in large numbers in the lymph nodes and spleen. They constitute in the adult from 20 to 25 per cent of the total number of white cells in blood and are the commonest cells found in lymph. In childhood lymphoid tissue is much more abundant than in adult life and the lymphocytes are more numerous. They amount to from 50 per cent or more of the leukocytes in early childhood and to about 35 per cent at the age of ten years.

(2) **LARGE LYMPHOCYTES** These resemble the preceding in general appearance but are considerably larger, being 12 or more microns in diameter. The cytoplasm forms a wider zone about the nucleus, which is oval or kidney shaped. These cells are found in insignificant numbers in adult blood but are more plentiful in the blood of young children. They are largely confined under physiological conditions to the lymphoid tissue, but even here they are greatly outnumbered by the small lymphocytes. They are considered by many as a younger form of the small lymphocyte.

(3) **THE MONOCYTES** are from 10 to 15 microns in diameter. They possess a relatively larger amount of cytoplasm. The nucleus has a deep indentation on one side, which gives it a kidney or saddle bag shape. On the supposition that this cell represented a stage in the development of the polymorphonuclear leukocyte, it was called the "transitional leukocyte" by Ehrlich. This view has since been shown to be wrong, for the monocyte bears a relationship to the lymphocytes rather than to the polymorphonuclears. It has been mentioned that the monocyte contains, like the lymphocytes, azurophil granules in the cytoplasm. The monocytes are actively motile and phagocytic, and are considered by most observers to be derived from fixed histiocytes (p. 105). Such an origin would class them as circulating elements of the reticulo-endothelial system. According to Maximow, however, they arise from lymphocytes. They constitute from 5 to 7 per cent of the white cells.

Small numbers (0.2 per cent) of a slightly different type of monocyte are also found in blood. Its nucleus

¹ The term leukocyte is employed by most authors to denote all the white cells, and this from the simple meaning of the word seems logical. Some, however, confine the term to the granulocytes. The first of these usages will be followed in this text.

instead of being kidney-shaped is round or oval. It was previously known as the large mononuclear leukocyte, but it is probably simply a younger form of the preceding variety.

II THE GRANULOCYTES

These are divided into three groups according to the staining reactions of their granules. One type—the *eosinophilic*—stains with acid dyes, e.g., eosin, another—the *basophilic*—stains with basic dyes, e.g., methylene blue, and the third type—the *neutrophilic*—with neutral dyes, i.e., mixtures of acid and basic dyes. These staining reactions apply to human leukocytes, but such distinctions cannot always be made in other animal species. The nucleus of a granulocyte is composed of two or more lobes connected together by strands of chromatin.

(1) The **EOSINOPHILS** are not numerous, they amount to no more than 2 to 4 per cent of the total white cell count. The granules which are oval and much coarser than those in the other two varieties are stained a bright red with eosin. The cell is also slightly larger and the nucleus usually bilobed. In certain pathological conditions which will be mentioned later they may form a much larger percentage of the leukocyte population.

(2) The **BASOPHILS** are present to the extent of only 0.15 per cent or less. Their granules stain deeply with methylene blue. Their significance is not known. They have been considered by some observers to be degenerated neutrophils, but there appears to be little doubt that they are a distinct type and like the other granulocytes a product of the bone-marrow. Support is lent to the latter view by the fact that they are increased in conditions associated with excessive marrow activity, e.g., chronic myelocytic leukemia and polycythemia vera. They are also increased in chronic inflammation of the accessory nasal sinuses.

(3) The **NEUTROPHILS** are by far the most numerous, constituting from 65 to 70 per cent or more of the total number of white cells. Their granules are quite small and are stained a violet tint with neutral dyes. As will be seen presently the neutrophils are actively ameboid in character, i.e., they are capable of locomotion and ingest foreign particulate matter. They are about 10 or 12 microns in diameter. Their nuclei show a variable number of lobes depending upon the age of the cell.

The Arneth count or index

It was pointed out by Arneth that the number of lobes in any neutrophil depends upon the cell's age,

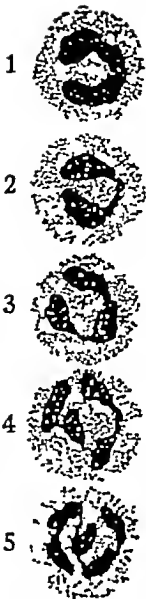


FIG 10 1 Arneth stages

the older cells having the larger number.² A five-lobed nucleus for instance indicates a stage in the life of a cell just preceding its final dissolution, an unlobed but deeply indented nucleus, a very young cell. Five stages in the life history of the polymorphonuclear leukocyte are therefore distinguished corresponding to the number of lobes developed in the nucleus. A count of the nuclear lobes in the cells of a blood film will give the proportion of cells of different relative ages. In figure 10 1, stage I shows a nucleus with a single lobe. Constriction of the nucleus can be seen but the nuclear substance is continuous from one part to the other. In stage II the nucleus is partitioned into two parts which are connected only by chromatin threads. In the next stage 3 lobes are seen and so on to the last or senile stage in which the nucleus has 5 or more lobes. The cells of this stage are large, edematous and non-motile, their granules strain poorly or not at all. In some conditions, e.g., acute septic infections and pernicious anemia, enormous neutrophils (up to 20μ) with a great number of nuclei are seen. These are known as *macropolycytes*.

The Arneth index is determined by counting the number of nuclear lobes in each of 100 neutrophils. The cells in the different stages are expressed as percentages of the total. The count under the ordinary conditions of health is as follows:

	Percentage of leukocytes
Stage I	5
Stage II	30
Stage III	45
Stage IV	18
Stage V	2

² The extent of the previous activity of the cell, rather than its age, may be the important factor determining the number of nuclear lobes.

TABLE 14

Showing Schilling index and differential count of mature cells, figures indicate percentages

NEUTROPHILS				Eosino- phils	Baso- phils	Lym- pho- cytes	Mono- cytes
Myelo- cytes	Juven- ile meta myelo- cytes	Older meta myelo- cytes	Nu- cleus lobular				
0	0-1	3-5	55-70	2-4	0-1	20-25	5-7

In certain diseases the youngest cells (Stage I) are much more numerous and may constitute 50 per cent of the total. There may be an entire absence of cells in the later stages (IV and V). An increase in the percentage of cells of the earlier stages is spoken of as a "shift to the left." It is seen in conditions which stimulate the bone marrow to a greater production of white cells, e.g., pyogenic infections. It is also seen in tuberculosis and after exposure to the X rays and after the injection of thyroid extract. In children a shift to the left occurs much more readily than in adults. In pernicious anemia the percentages of the older cells increase—"shift to the right"—and in some cases, as mentioned above, macropolycytes appear. Except in the case of the senile non motile cells a relationship between the phagocytic activity of a particular cell and its age has not been demonstrated.

The Schilling index employs a simpler classification of the neutrophils but includes marrow elements. Four stages are recognized (see table 14) (a) the myelocyte which shows a single spherical nucleus, (b) young metamyelocyte with a slight indentation of the nucleus, (c) older metamyelocytes with the process of lobulation definitely indicated, this is known as the band cell of Schilling or "Staff" cell and corresponds to the first stage of Arnet, (d) older neutrophils, i.e., the other stages of Arnet. Stages (a) and (b) are not found normally in the blood. They appear when a pronounced "shift to the left" occurs (See frontispiece and p 110).

The non motile cells of the last or fifth stage of Arnet appear, periodically in increased numbers—in "showers"—in the blood stream. They are replaced by young cells from the marrow. Like the red cell the dying neutrophils disintegrate in the circulation or are disposed of by the macrophages of the spleen or the tissues. The life span of the neutrophils has been variously estimated. By some it is believed to be no more than about 3 days and by others no more than a few hours. Ponder, however, puts it at 21 days. He induced a leukocytosis and shift to the left in the Arnet stages by the injection of thyroid extract and fol-

lowed the blood picture until the poly nuclear count returned to normal. Since the rise in the count is due to discharge of young cells (Stage I) from the marrow, when the count again showed the normal percentage of cells of Stage V it was assumed that the discharged cells had reached the end of their life span. At any rate the average life of the neutrophils is apparently much shorter than that of the red cell (p 74).

THE FUNCTIONS OF THE LEUKOCYTES

The neutrophilic polymorphonuclear leukocytes as well as the monocytes and other reticulo-endothelial elements constitute probably the most important elements which the body possesses for its defense against invading microorganisms. Their power to attack bacteria depends upon their motility and a proclivity for the ingestion of solid particles. The latter action, which was first demonstrated by Metchnikoff, is termed *phagocytosis* (phago—I eat). These two varieties of white blood cell are free lances among the body cells, they wander from place to place through the tissues and practically no part of the body is barred to them. They insinuate a process (*pseudopodium*), improvised at the moment from their cell protoplasm, through one of the joints in the endothelium of the capillary wall. Then by causing the semi fluid substance of the cell body to stream into the protoplasmic projection, they pass out of the blood vessels "at will." By this action of *diapedesis*,³ as it is called, myriads of white corpuscles may pass out of the vessels in a remarkably short time. Reaching a point where the bacteria have entered the body they surround the threatened area and proceed to destroy the invaders. If, for instance, an actively inflamed region should be examined under the microscope, masses of neutrophils would be seen, and many of these would be observed to hold bacteria imprisoned within their bodies. As many as 15 or 20 microorganisms may be seen at times within a single cell. It has been shown that the bacteria are ingested alive and remain so for a time within the leukocyte (fig 102).

When a tissue such as the mesentery or web of a frog, in which the capillaries are clearly visible, is examined in the living state a short time after a culture

³ The term diapedesis, literally a "leaping through", is sometimes applied to the passage of red cells through an unbroken capillary wall, but the term is scarcely appropriate for a passive process of this nature.

of bacteria has been injected into it, the small vessels leading to the site of inoculation are found swarming with neutrophils. In the tissues round about, the ameboid cells are seen moving somewhat ponderously hither and thither to engulf the offending bacteria. When the latter are intensely virulent in nature this normal leukocyte reaction may be seriously depressed. The monocytes, though much less numerous, also join in the general attack and show their phagocytic propensities to a marked degree. After the first flooding of the tissues with neutrophils and monocytes, numbers of the latter come to rest and together with other reticulo-endothelial elements of the tissues undergo transformation and aid in isolating the infected area from the neighboring healthy tissues. Until this is accomplished the danger of the infection becoming more wide-spread always exists. In their struggle against bacteria, equipped as these are with powerful toxins, many of the white cells are killed. These collect within the center of the inflamed area together with exuded plasma, liquefied tissue cells and a few red cells that have escaped through the injured walls of the capillaries. This material constitutes pus, and the so-called pus-cells are dead leukocytes. The circumscribing wall and its semi-fluid contents constitute an abscess. By the action of the phagocytes, aided by a protein-digesting ferment (protease) which they elaborate, the overlying structures whether connective tissue, mucosa or skin are in part removed piecemeal. In this way a communication with the exterior is effected and the contents of the cavity are discharged.

Not only bacteria but practically any foreign material, whether a rose-thorn or a catgut suture, is attacked and removed if possible, or loosened by the neutrophils aided by the monocytes and other phagocytic cells of the tissues. The removal of dead tissue or of blood clot or the separation of necrotic from living structures is accomplished in the same way. Devitalized bone though not removed in its entirety, unless it be of very small size is nevertheless eroded and separated from the living tissue by the leukocytes. The disappearance of effete organs such as the tail and gills of the metamorphosing tadpole or the creeping muscles of insect larvae, as these develop to the mature form, is effected by similar phagocytic cells. The application of heat to a part also attracts leukocytes in large numbers to the capillaries from which they immediately commence to migrate.

The activity of the leukocytes is best studied by the method of Sandison and Clarke, in which a transparent chamber is inserted into the tissues, e.g., the rabbit's ear. After a time fine vessels grow into the chamber through openings in its sides which may be examined under the microscope. Another very simple method is that of *supravital staining*. A thin film of a non-toxic (supravital) dye, e.g., neutral red, azure, or brilliant cresyl blue, is laid upon a glass slide and allowed to dry,

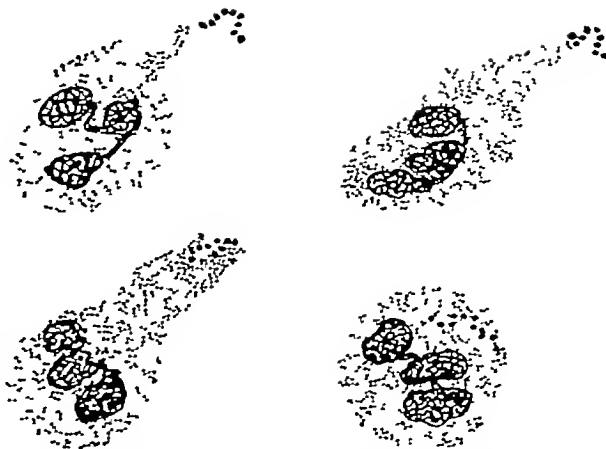


FIG 102 Drawing of a neutrophil at half-minute intervals to show motility and phagocytosis of bacteria

a film of blood is laid over this and covered with an ordinary cover glass which is then sealed with vaseline around the edges. The preparation is kept warm and examined under the microscope, the cells remaining alive and active. The neutrophils seen in such preparations are not uniformly spherical as in fixed smears, but are continually changing their shape. Pseudopodia are in constant movement and the granules can be seen streaming through the cytoplasm with each movement of the cell. The rate of progression of the neutrophil is from 30 to 35 microns per minute at body temperature (fig 102).

Of the functions of the other varieties of granulocytes—the *eosinophils* and *basophils*—little is known. The eosinophils are not markedly motile and only slightly phagocytic. Possibly the basophils, like the “mast” cells of the tissues, are a source of heparin (see also p 112).

The *lymphocytes*—though, generally speaking, they are not phagocytic—appear to exert such an action upon certain pathogenic microorganisms, notably, pneumococci types III and IV. A great migration of lymphocytes characterizes certain chronic types of inflammation.

An important function of the lymphocytes is the manufacture of serum globulin, both beta and gamma fractions having been found in extracts of lymphocytes. Immune substances (antibodies) are recognized as being associated with the gamma globulin fraction, so it was presumed that the lymphocytes occupied a key position in defense reactions of an immunological nature. The investigations of White and Dougherty have gone a long way to substantiate this idea. The lymphocytes and serum of mice which had been immunized to

sheep erythrocytes contained an antibody which lysed these cells, but none was found in the lymphocytes or serum of non-immunized mice. Furthermore, extracts of the lymphocytes of the immunized animals contained antibody in a concentration from six to eight times higher than that found in the serum. Antihemolysins to staphylococcus toxin have also been demonstrated in the lymphocytes of mice immunized to this toxin. The production of antibodies in the lymph nodes was first demonstrated by McMaster and Hudach in 1935, and Ehrlich found that a threefold increase in output of lymphocytes from the nodes accompanied the antibody production.

The manner in which the lymphocytes add globulin to the plasma appears to be by a process of cytoplasmic budding and ultimate dissolution in the lymphoid tissues and blood stream. Budding can be observed within a Sandison-Clarke chamber inserted into the tissues.

The supply of globulin to the blood is apparently under the control of the pituitary gland through the action of its adrenocorticotrophic hormone upon the adrenal cortex. The pituitary-adrenal influence was demonstrated by White and Dougherty in the following way. Rabbits were immunized to sheep corpuscles and, after the specific antibody had appeared in the circulation, were left untreated for a period of three months, when all antibody had disappeared. They were then divided into four groups. One group was injected with the original antigen, no significant rise in antibody concentration of the blood of these animals occurred. Of the remaining three groups, one was injected with steroid fractions of the adrenal cortex, another with the adrenocorticotrophic hormone of the pituitary (ACTH) and the third with an aqueous extract of adrenal cortex. The steroid hormones caused an increase in antibody concentration, nearly as great as that caused originally by immunization,—a maximal dissolution of lymphocytes in the lymphoid tissues and a scarcity of lymphocytes in the circulation. The response to the pituitary principle was also pronounced, though not as great as that given by the steroids, while the aqueous extract caused a relatively small reaction. The adrenal cortical fractions which exert this effect are those with an oxygen atom in the 11 position (see chapter 59). Corticosterone and 11-dehydro-17-hydroxycorticosterone are therefore effective in this respect but not desoxycorticosterone.

It has also been found that injections of adrenal cortical hormone, or of the pituitary principle, increase the bulk of lymphoid tissue as a result of edema, the edema fluid containing larger numbers of lymphocytes undergoing dissolution. Later, after the edema has subsided, the tissue shrinks until its weight is less than before the injections. Adrenalectomy is followed by an increase in the mass of lymphoid tissue.

The reduction in lymphocytes in response to adrenal cortical hormones and ACTH is of moderate degree. The eosinophils show a much more pronounced reduction in number following the administration of these principles. The eosinopenia occurs so consistently and is so delicate that it is employed as a means of assaying the potency of cortisone and ACTH. Adrenaline also causes eosinopenia, probably by stimulating ACTH production. Adrenalectomized animals show a lymphocytosis, and overgrowth of lymphoid tissue is not an uncommon feature of adrenal insufficiency (Addison's disease) in man.

THE FATE OF THE LYMPHOCYTES

This has been a puzzling question for a number of years, for it has long been known that immense numbers of these cells disappear from the circulation daily. According to one estimate, the entire population of lymphocytes in the circulation of the dog are replaced twice each day (Drinker and Yoffey), replacement 5 times daily is given for the cat (Sanders and associates). Some have thought that lymphocytes were transformed into other types of cell, others again that they returned to the lymph nodes and were destroyed in the germinal centers. But the most widely accepted explanation of the wastage of lymphocytes has been that they are shed from the mucosa of the gastro intestinal tract, and no doubt large numbers are discharged from the body in this way. Nevertheless, little effect is exerted upon the disappearance of lymphocytes in rabbits by the removal of the entire gastrointestinal tract and the injection of some 300 million leukocytes into such a preparation does not cause a rise in the leukocyte count. It appears now from the work of the authors mentioned above that the large daily turnover of these cells is to a great extent at least the result of their dissolution in the lymphoid tissues and blood, thus globulin is supplied to the plasma (see also chapter 11).

VARIATIONS IN THE NUMBER OF LEUKOCYTES IN THE BLOOD STREAM

LEUKOCYTOSIS

In the event of some damage to the tissues which calls forth a leukocytic response, not only is there a migration of leukocytes from the blood to the site of injury, but also a discharge of these cells from the marrow and an increase of their number in the general circulation. Instead of the normal count of seven or eight thousand per cubic millimeter the colorless cells may number 20 to 30 thousand within a short time. *Leukocytosis* is the term used to designate an increase in the total number of white cells. All varieties of the white cells do not necessarily share in the increase. In one instance it may be the neutrophils, in another the lymphocytes or the eosinophils that are increased, and it is the presence in abnormal numbers of one or other of these which then gives the high total leukocyte count. It is often of great diagnostic value to know which type of cell is responsible for the leukocytosis, and in order to determine this a so-called *differential count* of the cells is made. That is, the numbers of the different types in a stained smear of blood are counted and their percentages of the total count determined. Also, changes in the proportions of the different white cell types may occur though their total number be normal. Such alterations can be revealed only by a differential count. The following example is given in illustration, the lymphocytes and monocytes are relatively increased, the neutrophils and eosinophils reduced.

Total white cells per cubic millimeter	7500
Lymphocytes, per cent	35
Monocytes, per cent	8
Neutrophils, per cent	56
Eosinophils, per cent	0 7

Very commonly an increase in the neutrophils is entirely responsible for the leukocytosis. On this account the latter term is used frequently but somewhat loosely to imply an increase in the count caused by the neutrophilic elements alone. *Neutrophilia* is a more precise term that has come into use for the latter condition. *Lymphocytosis*, *monocytosis* and *eosinophilia* are the respective terms employed to designate increase in the other elements.

Acute infections by the pus-forming organisms—staphylococcus, streptococcus, etc., are the most potent causes of a neutrophilic increase. On this

account the examination of the white cells furnishes a valuable diagnostic sign for the detection of hidden inflammatory conditions, e.g., appendicitis, empyema, etc. A neutrophilic leukocytosis occurs also in pneumonia, whooping cough, scarlet fever and some other infectious fevers. The nuclear count shows an increase in the young stages at the expense of the older (p. 93).

Pteroylglutamic ("folic") acid is a dietary constituent required by many species for the production of granulocytes by the bone marrow, and for the maintenance of the normal population in the circulation. Absence of this vitamin from the diet of monkeys and chicks is followed by a profound leukopenia.

Chemotaxis, chemical factors in inflammation

Chemotaxis is the term applied to the unknown "force" which draws the white cells from the blood stream, to the point of injury in the tissues. This property of responding by a locomotory movement to chemical substances is not peculiar to leukocytes, but is possessed by many types of free-living, unicellular organisms. Chemotaxis may be positive (i.e., attractive) or negative (repellant) and, though the former is most commonly observed in leukocytic behavior, certain chemicals produce the opposite effect.

It was formerly thought that the chemical properties of the bacterial toxin were responsible for this effect, but it has since been shown that nucleic acid and its derivatives guanine, adenine, adenosine, etc. or some other principle supplied by the tissues (see below) are the specific stimulants. Injections of these substances cause a rapid rise in the leukocyte count. Toxins or other injurious agents act probably indirectly by liberating nucleic acid from the tissue cells as well as from injured leukocytes themselves. It has been suggested that the actual force which attracts the cells from the vessels to the tissue focus may have changes in surface tension of the blood cell membrane as its basis, for during the early stages of the inflammatory reaction the leukocytes in the small vessels near the injured site appear "sticky." They collect and cling to the walls of the vessels and are thus separated from the red cells which occupy the axis of the stream. An artificial cell model may be used to illustrate the surface tension theory. When a small globule of mercury is placed in a weak solution of nitric acid, the globule moves rapidly toward a crystal of potassium dichromate dropped near it upon the surface of the solution, as a result of a chemical reaction leading to surface tension effects, or if the mercury is placed in a dilute dichromate solution, it will move away from a drop of nitric acid in its vicinity (negative chemotaxis). Experiments, however, upon unicellular organisms such as the amoeba, to which the white cell bears a strong resemblance, have failed to show that the spontaneous movements are surface

tension phenomena. It must be admitted that the problem of chemotaxis remains unsolved. Phagocytosis does not necessarily depend upon chemotaxis but may occur quite apart from it, as when chemically inert material, e.g., particles of carbon or silica are engulfed by leukocytes coming into contact with them at random. On the other hand, a positive chemotactic effect may not be followed by phagocytosis, for neutrophils may be attracted by some dead or foreign material and move toward it with the apparent "intention" of devouring it, but not do so.

Menken has obtained a nitrogenous, crystalline principle from inflammatory exudates which increases capillary permeability, allowing the free escape of plasma protein, and induces the migration of leukocytes through the capillary wall. This substance, named *leukotaxine*, appears to be a simple polypeptide and is related neither to histamine nor to nucleic acid. It has also been found by others in the succus entericus of the rabbit. It seems to be the factor responsible for the swarm of leukocytes which in infective and inflammatory states are attracted from the circulation into the tissues of the affected part. Four other substances have been recovered by Menken from inflammatory exudates, called the *leukocytosis promoting factor* (L P F), *necrosis*, *pyrexin* and the *leukopenic factor*, respectively.

The *leukocytosis promoting factor* causes the discharge of immature leukocytes from the bone marrow and hyperplasia of granulocytes and megakaryocytes within the latter. It is a pseudoglobulin, or closely associated with this fraction of the exudate, and is presumably responsible for the leukocytosis which is so often a feature of the blood in infective states. Though increased manufacture of granulocytes is induced by L P F, the rapid increase in leukocytes suggests that preformed cells are discharged into the circulation, and there is evidence that in the early stages of leukocytosis capillary sinuses of the marrow, harboring masses of leukocytes, are suddenly opened up and their cellular contents discharged. This suggests a vasomotor reaction, which is substantiated by the fact that even saline (1 cc.) injected intravenously into a rabbit causes about 300,000,000 mature neutrophils to enter the general blood stream within 1 hour.

Necrosis, which is associated with the euglobulin fraction of inflammatory exudates, is the substance which causes tissue injury, lymphatic obstruction and necrosis in inflammation. The lymphatic blockage, according to Menken, is salutary since it tends to limit the spread of the infection. Injection of this substance intravenously into animals causes widespread injury, e.g., hemorrhages into the gastro-intestinal tract, focal necrosis of the liver and leukocytic infiltration of the kidneys. *Necrosis* is thermolabile and nonpyrogenic.

Pyrexin is a thermostable fraction associated also with the euglobulin. It appears to be a glycoprotein. It induces fever.

The *leukopenic factor* causes leukopenia as a result of the trapping of leukocytes in the lungs, spleen and liver. This factor causes nausea and vomiting when injected intravenously. It is closely associated with pyrexin and, though it can be separated by incomplete hydrolysis from the pyrogenic factor, it has not been shown with certainty that the two factors are separate and distinct.

Inasmuch as the leukocytes, especially the neutrophils, are essential elements in the defensive mechanisms of the body against infective microorganisms, their attraction to an infected part must be looked upon as a physiological and salutary response. It should, therefore, not be discouraged in any way by agents, e.g., many antiseptics, which, though themselves inimical to bacterial growth, defeat their own purpose by destroying the leukocytes, or reducing their activity. Such agents may even act to repel the neutrophils from the injured region. Sulfanilamide, on the contrary, is claimed to actually stimulate leukocyte activity either directly or by rendering the invading microorganisms less resistant, or more "appetizing", to the phagocytes. Some interesting observations have been made by Mallery and McCutcheon on the movements of leukocytes in attack, which give a meaning to the well worn phrase "lowered resistance". The neutrophils in samples of blood from patients acutely ill and from those convalescent from various diseases were observed, and their rates of approach to a minute clump of bacteria measured and compared with the rates of approach of the observer's cells under identical conditions. In the acutely ill patients the rate of approach was 9.7 microns per minute, as compared with the normal of 16.1 microns per minute. No significant difference was observed between the phagocytes of the convalescent patients and those of the observer.

Physiological leukocytoses. It had formerly been taught that an increase in the neutrophils occurred during digestion—*digestive leukocytosis*—but it seems that this was a misconception. These cells show spontaneous rhythmical variations in their numbers, the total white cell count reaching its maximum of about 7000 to 8000 in the afternoon, and its minimum, 5000 to 6000, in the early morning. These variations occur quite independently of meals. Leukocytosis also occurs during pregnancy, parturition and menstruation in muscular exercise and after adrenaline administration, or in states such as fear, pain, anoxia, etc., which cause

the liberation of adrenaline from the adrenal gland. In infants and young children the leukocyte count is considerably higher than in adults, the count is also less constant in infancy, varying without apparent cause by two thousand or more per cubic millimeter.

Eosinophilia, or increase in the number of circulating eosinophils, occurs in several conditions, notably allergic states, e.g., asthma and anaphylactic shock, and in infections by various animal parasites, e.g., *hook-worm* (*ankylostoma duodenale*) disease, in which the eosinophils may be 30 per cent of the total white cell count, and *trichinosis*. In the latter infection there is a general leukocytosis, with the eosinophils running as high as 50 per cent of the total. Infections with hydatids, ascari and other worms also cause eosinophilia to a greater or less degree. The significance of this association of eosinophilia with parasitic infection is unknown. These white cells are also increased in Loeffler's disease and in a number of skin diseases, the tissues in the neighborhood of the cutaneous lesions may be infiltrated with eosinophils. During the acute stage of pyogenic infections the eosinophils are usually reduced in number (eosinopenia), in the convalescent stage they are, as a rule, increased. The eosinopenia caused by ACTH and by adrenaline has been mentioned.

Lymphocytosis. The neutrophils are not stimulated by tuberculous, malarial, or syphilitic infection. In the active stage of such conditions either an absolute or relative increase in the number of circulating lymphocytes is the rule. In other chronic inflammatory states and in infections with the colon or diphtheria bacillus also, it is the lymphocytes rather than the neutrophils that are increased in number. They indicate in general an inflammatory condition that is undergoing repair, is being held in check, or at the most is making slow progress. Lymphocytosis, therefore occurs as an aftermath of acute infections. The neutrophils on the other hand represent the "shock troops" and their presence indicates that a more active campaign is being waged. In young children a relative lymphocytosis is the rule.

Monocytosis. An increase in the monocytes is seen much less frequently than that of the other leukocytic types. Apart from monocytic leukemia and glandular fever (see below), the chief conditions in which they appear in greater numbers than normal are tuberculosis, malaria, syphilis, brucellosis and bacterial endocarditis. According to Cunningham, a decline in the lymphocyte count with an increase in the monocytes in pulmonary tuberculosis is an indication that the tuberculosis process is being arrested.

PATHOLOGICAL INCREASES IN THE LEUKOCYTE POPULATION. LEUKEMIA, GLANDULAR FEVER

The leukocytic increases discussed in the foregoing sections, even those associated with abnormal states, are moderate in degree, and are due to reactions which in themselves are of a "pur-

poseful" character and on the whole physiological. But in the *leukemias* we find an altogether uncontrolled and often relatively enormous increase (up to 1,000,000 per cubic millimeter) in the number of leukocytes, and a distorted mode of white cell production with the appearance of immature forms in the circulation.

Leukemia is now generally looked upon as a malignant neoplastic disease of the white corpuscles, as with other malignancies, its cause is unknown.

The disease occurs spontaneously in the fowl (*chicken leukosis*) and in many mammalian species, and can be transmitted by the transfusion of viable leukemic cells of one animal into the circulation of a normal animal. The transfused cells multiply in the blood of the host. That leukemia can be transferred in a similar way from man to animals has not been proved. The disease in fowl can also be transmitted by cell-free blood passed through a Birkefeld filter, and is therefore believed to be due to a virus of some sort, probably a chemical substance. This substance does not pass from an affected to a normal bird—under natural conditions. The disease in man, or in other mammals, does not appear to be due to a transmissible virus. Whatever may instigate the abnormal blood state, its influence is impressed upon the earliest progenitors of the white cells in the bone marrow or lymphoid tissue, which results in the production of abnormally formed and malignant blood cells. This same morbid state may produce tumor-masses (lymphosarcoma or myelosarcoma), clearly indicating its malignant nature.

In experimental leukemia in mice, which is closely similar or identical with the human disease, genetic factors are prominent. The influence of heredity is often evident also in human leukemia. By inbreeding mice which are susceptible to the spontaneous development of leukemia, strains have been produced in which from 80–90 per cent of the animals show the disease at the age of from 6 to 9 months. Little is known of the other factors which may be concerned, whether hormonal, humoral, metabolic or nutritional (e.g., vitamin or mineral deficiency). Miller and Turner and Wearn and his colleagues have found that the urine of patients with leukemia contains principles which stimulate leukopoiesis in guinea pigs and produce a blood picture in these animals resembling that of human leukemia. The actions of these substances suggest that they are specific stimulators of leukocyte production. It is postulated

that they are normal principles (one of which induces hyperplasia of lymphocytes, the other of granulocytes), but produced in excess in leukemia. The lymphocyte-stimulating principle appears to be a hydroxy-acid, while that which causes granulocyte hyperplasia is believed to be a non-carbonol-acid. A second myeloid-stimulating substance, obtained from the urine of patients with myeloid leukemia, has the character of a protein or of a glycoprotein.

It has been suggested by Miller and Turner that these two principles normally act reciprocally to control leukopoiesis and maintain a nice balance between the production of myeloid and lymphoid elements. The myeloid stimulating substance induces proliferation of myeloid elements, inhibition of lymphocytic hyperplasia and maturation of lymphocytes, while the lymphoid material acts to stimulate the proliferation of lymphocytes, to inhibit myeloid proliferation and to induce myeloid maturation.

The incidence of the disease in mice is greater in females than in males, which has suggested an influence of the sex hormones, and the administration of estrogen or androgen has been found to increase or reduce, respectively, the susceptibility of these animals to the experimental transmission of the disease. 11-dehydro-17-hydroxycorticosterone for a time reduces the number of leukocytes in the leukemia of mice and causes the shrinkage of lymphosarcomatous tissue, little lasting effect upon the course of the disease has been observed. Adrenalectomy has a pronounced effect in increasing the susceptibility of mice to the experimental disease.

The thymus appears also to have a definite effect upon the incidence of spontaneous leukemia in mice, removal of this structure causing a very marked reduction. The nature of the relationship is unknown.

The leukemias are usually classified on the basis of the cell type—lymphoid or myeloid (e.g., granulocytic)—involved in the malignant hyperplasia. Thus, *lymphatic* (*lymphoblastic* or *lymphocytic*) *leukemia* and *myeloid* (*myelocytic* or *myeloblastic*) *leukemia*, respectively, are usual designations. Either of these may be classed as *acute* or *chronic*. This latter division is based not so much upon the clinical course of the disease as upon the degree of immaturity of the cells in the blood and bone marrow, the acute form, as compared with the chronic, being characterized by cells in the earliest

stage of development. In *acute myeloid* (*myeloblastic*) leukemia the type cell is the *myeloblast*. The myeloblast is an entirely abnormal cell and not simply an immature leukocyte, as seen normally in the bone marrow. It is a large cell about 20 microns in diameter with a single round or oval nucleus which nearly fills the cell. The chromatin is distributed evenly throughout the nucleus with little condensation into masses. The nucleus contains from 4 to 5 nucleoli. The cytoplasm is strongly basophilic, shows no granulation and is thin and irregular at the cell boundary. In suitable preparations a slow snail like movement of these cells can be demonstrated by slow cinematography. Undifferentiated myelocytes are also found in the blood. In the chronic form of *myeloid leukemia* (*myelocytic leukemia*) myeloblasts are infrequent and the myelocytes more plentiful.

In *acute lymphatic* (*lymphoblastic*) leukemia, the characteristic cell is the *lymphoblast*. This cell closely resembles, but can be distinguished from the myeloblast by the coarser chromatin structure in its nucleus, by possessing only one or two nucleoli, and by showing a characteristic movement described by Wintrobe as "stately" and apparently purposeful. In the chronic form of *lymphatic* (*lymphocytic*) *leukemia*, the leukocytosis is due to small lymphocytes, which make up over 90 per cent of the total number of white cells, and may be 250,000 per cubic millimeter. Few immature cells are seen.

Leukemias in which the type cell is the monocyte, basophil, or even the megakaryocyte also occur rarely.

The somewhat redundant or contradictory terms *leukemic leukemia*, *subleukemic leukemia* and *aleukemic leukemia* are used, respectively, to designate the abundance, scarcity or absence of leukocytes in the blood. The first mentioned term is applied to those leukemias in which the blood picture is dominated by leukocytes. In subacute leukemia, the white cell count is normal or near the normal and only an odd immature form can be found, and in the aleukemic type few or no abnormal cells are present, the total count is not increased and may be subnormal. But this is only a phase of the disease, a rise in the leukocyte count eventually occurs, the leukemia being usually of the lymphatic type.

In the treatment of the leukemias, the destructive action of the X-rays upon lymphoid tissue and bone marrow is widely employed. The life of

sufferers from the chronic form is thereby prolonged, but neither X-ray nor any other agent is of any avail in the acute forms. Internal radiation with radioactive phosphorus (P^{32}) is also used in the chronic form and acts like X-rays, but has a more selective action. Within recent years several chemical agents have been advocated for the treatment of chronic leukemias, e.g., *urethane* (ethyl carbamate), especially in the chronic myeloid type, *nitrogen mustards*, 4-amino-methyl-pteroylglutamic acid (aminopterin) or other folic acid antagonist.

In *Hodgkin's disease*, which is allied to the leukemias, there is general enlargement of the lymphoid tissue and usually a moderate leukocytosis (15,000 to 25,000 per cu mm), in which neutrophils predominate and lymphocytes are reduced.

The monocytes and lymphocytes are increased, in the condition originally named *glandular fever* by Pfeiffer (1889) and *infectious mononucleosis* by Sprunt and Evans (1920). There is enlargement of the cervical lymph glands, spleen and liver, and a leukocytosis, usually not exceeding 20,000, of which the mononuclear leukocytes (monocytes and leukocytes) constitute from 60 to 90 per cent or more. A characteristic serological feature of this disease is the usual, though not invariable, finding of a high titer of agglutinins against sheep corpuscles (heterophil antibodies), as first shown by Paul and Brunnell.

LEUKOPENIA

Leukopenia means a reduction in the number of circulating leukocytes. It is seen in certain diseases, notably typhoid fever, and may be induced experimentally by injections of the toxin of the typhoid bacillus—or emulsions of the dead organisms, and also by the injection of Menken's leukopenic factor (p 98), the action of the adrenal cortical hormones and ACTH in producing a reduction in lymphocytes (*lymphopenia*) has been mentioned (p 96). Leukopenia is also a feature of "folic acid" deficiency in some species, e.g., monkey and chick. In some cases in which the white cells are reduced in number in the blood, the reduction is due to their attraction to some solid organ such as the lung or spleen. This has been shown by taking blood counts from various regions. In other words leukopenia may be due to a redistribution of leukocytes in the body, rather than to an actual reduction in their number. A temporary fall in the leukocyte count may precede a leukocytosis. Cer-

tain poisons, e.g., benzol, cause leukopenia by depressing the activity of the bone marrow (see also p 90).

Granulocytopenia, agranulocytosis, etc. *Granulocytopenia* is the term applied to an abnormally low leukocyte count due to the reduction in granulocytes. The lymphocytes and monocytes are but slightly reduced or not at all, so that their proportion of the total count is increased. One or both of these types of agranular cells sometimes show an *absolute* increase. There may be complete absence of granulocytes when the term *agranulocytosis* is applicable. In most instances the absence of granulocytes is associated with a severe septic or necrotic condition of the throat. This condition, called *agranulocytic angina*, is fatal in the great majority of cases. The cause of these states is unknown but the fault is evidently one of the bone marrow, and is probably induced by some toxic agent. The marrow shows, frequently, an almost complete suppression of granulocyte formation but is normal so far as erythropoiesis is concerned. In animals reduction in the granulocytes is readily induced by the administration of benzol, which acts specifically to depress marrow activity, and there is a belief that in some cases agranulocytosis is induced by certain benzol derivatives employed for their antipyretic or analgesic properties. The arsenobenzenes, dinitrophenol and, though rarely, sulfanilamide and sulfathiazole have been incriminated.

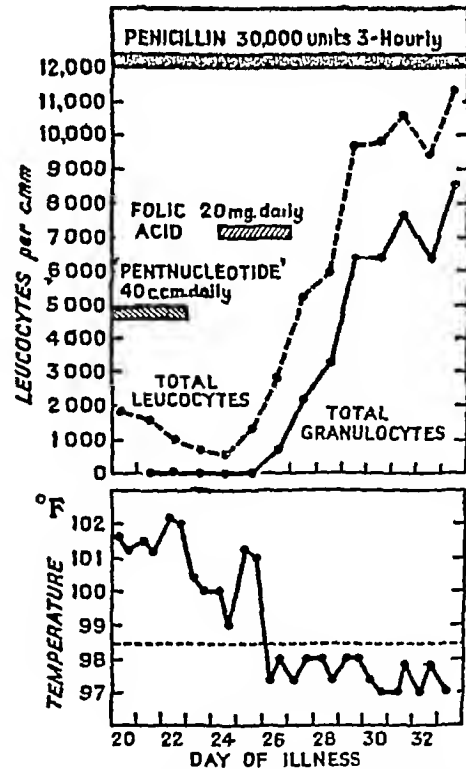


Fig 10-3 Showing the effect of folic acid upon the granulocytes, in agranulocytosis (After Black and Stanbury, slightly modified)

Amidopyrine was the first drug shown to exert this effect but other drugs, e.g., sulfonamides, arsenicals, certain derivatives of quinine, and chloramphenicol sometimes cause the disease. In the treatment of these conditions, pentose nucleotide has been employed with the object of stimulating the granulopoietic functions of the marrow, but pteroylglutamic acid ("folic acid") appears to be the most promising agent yet discovered for the treatment of granulocytopenia (see fig. 10.3).

In some cases of agranulocytosis the bone marrow shows an increase in the number of primitive cells of the granulocyte series (myelocytes and myeloblasts), and cells of these stages appear in the circulation. The maturation of granulocytes is apparently arrested at an early stage (p. 110). From the analogy between this fault in granulopoiesis and the erythropoietic abnormality seen in pernicious anemia, the condition has been termed *pernicious leukopenia*, or the *maturation type* of the disease.

THE BLOOD PLATELETS (THROMBOCYTES)

These are commonly stated to be simply fragments of protoplasm (i.e., non-nucleated) derived from the cytoplasm of the megakaryocytes (p. 112). Their colorless cytoplasm contains two types of granules. Those of one type are arranged centrally in clumps or chains and stain supravitaly with neutral red or azure blue. Those of the other type are discrete, stain supravitaly with Janus green and are scattered throughout the body of the cell. The platelets have an average diameter about a third that of a red cell, namely, 2.5 microns, and number from 200,000 to 400,000 per cubic millimeter. The most usual figure found in health is around 250,000 per c. mm. These blood elements vary considerably in shape. Their best known function is concerned with the mechanism of blood clotting (chapter 12). The disintegration of the platelets is said to occur more readily in blood drawn during the digestion of a meal of meat. The major portion of the histamine of the blood is contained in the platelets. They possess insignificant amounts of thromboplastin (see p. 114) but furnish a factor which accelerates the

first as well as the second stage of the clotting process.

Variations in the number of platelets occur in the following conditions. They are *increased* after a meal of meat, after hemorrhage, and in certain allergic conditions, in myeloid leukemia and in convalescence from infections.

They are *diminished* in purpura hemorrhagica, aplastic anemia, pernicious anemia, in anaphylaxis and in the acute stage of septic infections.

The number of platelets per cubic millimeter is determined most conveniently by diluting a sample of blood with a fluid composed of sodium citrate 3.8 per cent, formalin 0.2 per cent and brilliant cresyl blue 0.1 per cent, and counting immediately. The proportion of platelets to red cells (normally about 1 to 20) is determined. If the number of red cells per cubic millimeter be known then the corresponding number of platelets is readily calculated.

Besides their well-known rôle in the coagulation of blood (ch. 12) the platelets probably serve other functions. They have a pronounced tendency to agglutinate into masses and to form deposits upon any roughened surface or foreign material. Particles of India ink or microorganisms injected into the body become surrounded by a mass of agglutinated platelets. They may therefore aid in the body's defense against infective agents. It is probable also that they serve to seal leaks in the capillaries by adhering to small defects which may occur from time to time in the delicate endothelial wall. They constitute the first defense against the loss of blood from larger vessels. Collecting around the margins of the vascular wound they help to close it, or at any rate serve to fasten the clot, which subsequently forms, to the vascular wall, and through their action in inducing retraction and consolidation of the clot, to narrow the opening in the vessel and form a firm plug within its lumen. They liberate a vasoconstrictor principle (p. 115) which acts locally, and also a hypotensive material. The length of life of the platelets has been estimated at from 3 to 5 days.

CHAPTER 11

THE ORIGIN OF THE BLOOD CELLS—HEMATOPOIESIS

INTRODUCTION

There are two main schools of thought concerning the development of the blood cells in postnatal life (a) The *unitarian* or *monophyletic* school holds the view that all types of blood cell are derived from a common primitive free cell which they term the "*stem cell*" or *hemocytoblast* (b) The *dualistic* school believes in the existence of two distinct types of stem cell, one in the bone marrow which gives rise to the myeloid elements—erythrocytes, granulocytes and megakaryocytes—and the other in lymphoid tissue which is responsible solely for the genesis of the lymphocyte. It is not possible to speak unreservedly for either theory, though the unitarian view seems to have the balance of evidence in its favor and will be followed in this text. Both schools are agreed, however, that in the early embryo the mesenchyme gives rise to a primitive free cell from which all the blood cells are derived. It is also generally conceded that certain cells (reticular cells) of the bone-marrow and lymphoid tissues of the adult (p 108) are the representatives of the mesenchyme cells of the embryo and to such cells all the blood cells trace their lineage.

The point at issue is, "Do the bone-marrow and lymphoid tissues give rise to two cell types, with their potentialities restricted, the one to the development of myeloid elements, the other to the development of lymphocytes?" Or, "Do both types of tissue give rise to a primitive free cell—the stem cell or hemocytoblast—with potentialities for the production of all types of blood cells, but whose development along one or other line is determined simply by its immediate environment?" The latter is the monophyletic view (Pappenheim, Maximow).

It is not maintained, however, that the individual blood cells arise, under ordinary circumstances, in direct line of development from the stem cells, that is, each blood cell from a hemocytoblast. In health, the blood cells are produced through the multiplication of cells belonging to later stages of hemopoiesis (erythroblasts and myelocytes, p 111 and p 110). In other words a single stem cell is the ancestor of many millions of mature blood cells.

This type of development, involving the proliferation of cells of later stages and the production of daughter cells of the same type which then undergo maturation, is called *homoplastic hemopoiesis*.

Under pathological conditions, on the other hand, the stem cells may undergo active proliferation, and produce, directly, immature blood cells (erythroblasts of various ages, myeloblasts and myelocytes). This is termed *heteroplastic hemopoiesis*.

The scheme, p 104, and figures 11 4 and 11 5 will enable the reader to follow the description of the blood cell origins which will now be given.

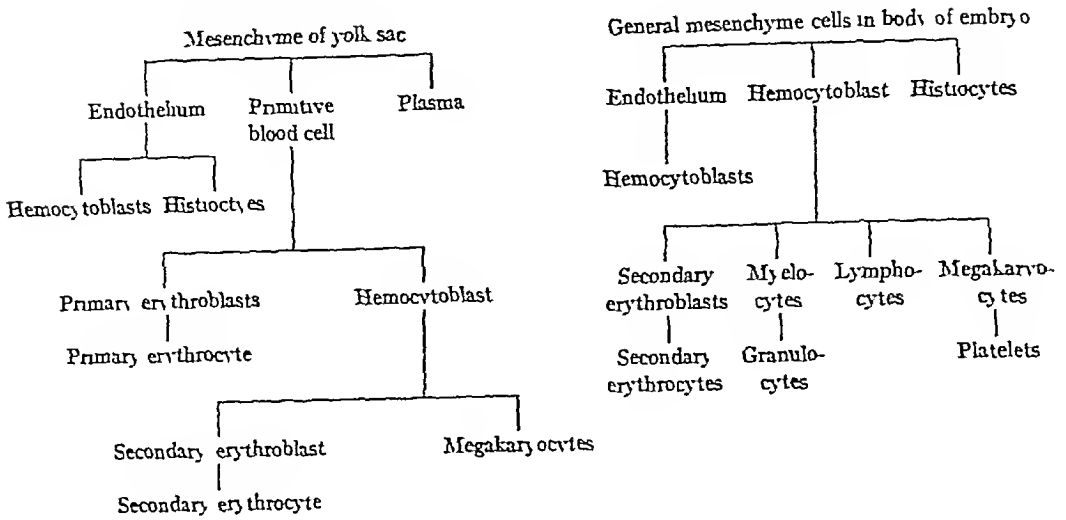
HEMATOPOIESIS (BLOOD FORMATION) AT DIFFERENT EMBRYONIC AGES

IN THE EARLY EMBRYO

THE YOLK SAC The first signs of blood and blood vessels appear in the mesenchyme of the wall of the yolk sac, i.e., outside the embryonic area proper. At a very early stage groups of mesenchyme cells in this situation—the *blood islands of Pander*—are first observed to arrange themselves into cords or columns which soon separate into a central and two outer layers.

The outer two enclosing layers form the walls of the primitive blood vessel—*primitive endothelium*. Of the central group some go to form the first blood cells, while the secretion or actual solution of others forms the plasma in which the cells are suspended. These free elements are known as the *primitive blood cells*. After this the mesenchyme loses its power to produce blood cells directly. A few primitive blood cells may arise from the primitive endothelium from time to time, but they are few in number and the process is believed to be relatively unimportant in most mammals, and does not persist for long. The further development of the primitive blood cells follows one or another of two courses (a) Some of the cells—*primary erythroblasts* and *erythrocytes*—acquire hemoglobin and serve as oxygen carriers. These are short lived and soon disappear forever. (b) The rest remain colorless and apparently unchanged from their primitive state, *they are very similar to, if not identical in appearance with, the large lymphocyte*. These cells are found in adult bone marrow and lymphoid tissue, and in small numbers in circulating blood. They are termed by Maximow "*hemocytoblasts*" and upon them the monophyletic conception of blood formation in post-embryonic life is based. They

DIAGRAM TO SHOW THE DEVELOPMENT OF THE BLOOD CELLS IN THE EARLY EMBRYO



are the 'stem cells' and are, according to the unitarians, potentially capable of producing any of the blood cells in the adult. Though these cells are identical, morphologically, with the primitive blood cells of which they are an older stage, apparently they are functionally different, for they form *secondary* erythroblasts and erythrocytes (which the primitive cells do not) but never *primary* erythrocytes (see chart, above). They also give rise to megakaryocytes. The latter are enormous cells (40μ) with multi-lobed nuclei. The primitive endothelium as well as the primitive blood cells gives rise to hemocytoblasts and also to a few histiocytes (p. 106) which show phagocytic proclivities, devouring degenerated red cells. In mammals few granulocytes are formed within the yolk vessels. They arise extra-vascularly from hemocytoblasts derived from the mesenchyme cells.

IN THE BODY OF THE EMBRYO While these stages of blood-development are proceeding in the mesenchyme of the yolk-sac, the heart and vessels are developing in the embryonic area. Soon the embryonic and extra-embryonic systems of vessels form communications with one another and the primitive blood-plasma, primary erythrocytes and hemocytoblasts flow into the body of the embryo. The mesenchyme cells of the general connective tissues of the embryo's body also form hemocytoblasts at this time. From these stem cells secondary erythrocytes and later, granulocytes and lymphocytes are produced. Blood formation throughout the general mesenchyme is, however, of short duration. Hemopoiesis soon becomes localized in the liver, spleen, bone marrow and lymph glands. Normally it is only in these situations that the mesenchyme cells exhibit their powers of producing hemocytoblasts, therefore, it is exclusively in these tissues that the hemopoietic function for a time is carried on.

IN LATE EMBRYONIC AND POST-NATAL LIFE

In the later part of pre-natal life of most animals, the liver and spleen (except the lymphoid tissue of the latter) lose the power to produce stem cells and so no longer serve as blood forming organs. The hemopoietic function from now on resides solely in the bone marrow and lymphoid tissues (p. 108). The marrow is concerned with the production of red cells, granulocytes and platelets, the lymphoid tissue of lymph glands, of Peyer's patches of the intestine, and of the spleen and thymus form lymphocytes. In certain animals, such as the opossum and frog, the formation of red cells (erythropoiesis) and of granulocytes (granulopoiesis) is continued throughout adult life by the spleen. In the bird, though the marrow is the chief organ for blood formation, the liver still retains in part its embryonic hemopoietic function.

In late embryonic and post-natal life the mesenchyme gives rise to three main types of cells

(a) Those which retain their embryonic potencies throughout adult life, being capable of producing hemocytoblasts (stem cells) and so of generating any or all of the blood cells. Some of these, called the *embryonic reticular cells*, are situated in the reticulum of the bone marrow and of the lymphoid tissue. Ordinarily these cells, as stated on page 111, are restricted in their hemopoietic activities, the great majority of the blood cells in conditions of health being formed by the divisions and re-divisions of cells of their own kind (homoplastic hemopoiesis). Cells having similar potencies are also present in the general connective tissues, and in these situations are spoken of as *undifferentiated*

mesenchyme cells Under ordinary circumstances these latter cells do not give rise to stem cells. As the result of some abnormal stimulus, however, their dormant powers inherited from their mesenchyme ancestry may become aroused and they may then give rise to the various types of blood cells. For example, areas resembling marrow tissue may be produced by the experimental stimulation of these undifferentiated cells, the several types of blood cells are formed within such areas. Maximow rendered the kidney of the rabbit necrotic by tying the renal vessels, and thereby induced in this situation hemopoietic activity resembling that which occurs in adult red marrow. Marrow-like tissue producing erythrocytes and granulocytes may also arise as a result of an abnormal stimulus in the spleen, liver, adrenals, aorta, lymph nodes and other sites. In certain forms of anemia, especially of infants, the extra-medullary formation of red cells sometimes occurs (in kidney, spleen, liver, etc.) and in leukemia, granulocytes, which normally arise only in the bone marrow, are produced by the spleen.

(b) The second type of cell has retained a certain measure of its embryonic characteristics. These cells are endowed with the remarkable power of altering their form and functions under appropriate stimulation. They are found in the general connective tissues lying among the fibroblasts. They are allied to the connective tissues on the one hand and on the other to certain white cells (monocytes) of the blood and so form a connecting link between the tissues and the circulating cells.

The cells of this group constitute what has been termed by Aschoff the *reticulo-endothelial system* (see below.)

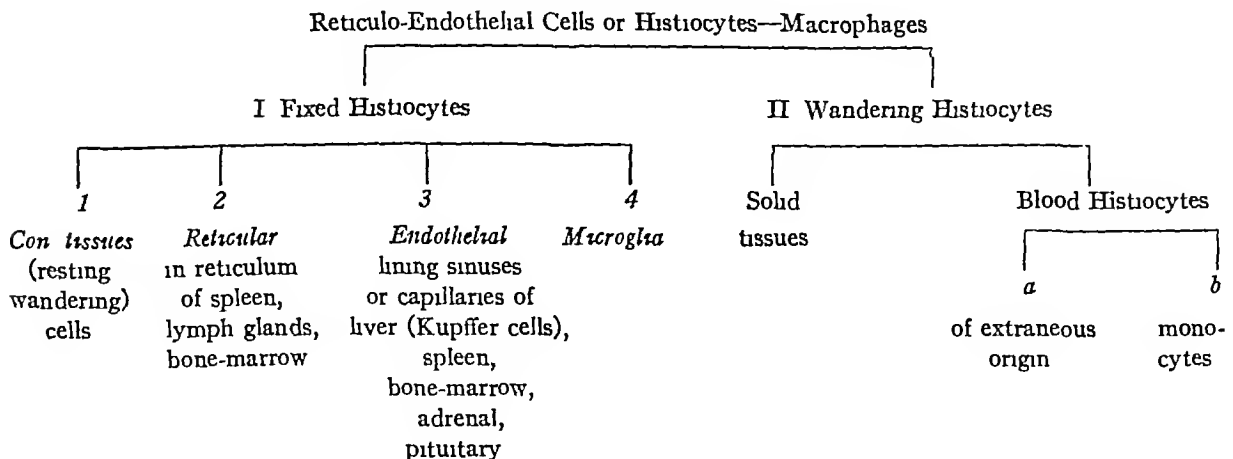
(c) In the general connective tissues the great majority of the original mesenchyme cells become transformed into the ordinary and completely differentiated connective tissue elements—the fibroblasts. These, once formed, remain practically unchanged in structure and in function.

THE RETICULO-ENDOTHELIAL SYSTEM

To the cells of this system the general term *histiocyte* has been applied. The term simply means a tissue-cell, and on this account is without descriptive value. But it would be difficult to coin a word that would embrace all the various cell-types of this system, and yet would be sufficiently explicit to distinguish them from some of the blood cells. *Pyrrhol* cells is a term that has been used in the past for a reason that will appear presently. The different types of histiocytes possess one characteristic in common. They are phagocytic for foreign particles of all sorts, and for this reason they were called *macrophages* by Metchnikoff to distinguish them from the much smaller phagocytes of the blood—the neutrophils or *microphages*. In one particular, however, the histiocytes differ from all other cells of the body—neutrophils included. They are stained in the living state by weak solutions of certain colloidal dyes—pyrrhol-blue, trypan blue, lithium carmine, etc. The vital or supravital staining reaction is simply a process of ultramicroscopic phagocytosis. That is, the fine ultramicroscopic particles of the dye are taken up from the solution (which is too dilute to stain ordinary cells) and as a result of their accumulation into larger masses in the cytoplasm become visible under the microscope. It is by means of this reaction that the macrophages are able to be detected among the ordinary tissue cells from which it is sometimes difficult otherwise to distinguish them.

VARIETIES OF RETICULO-ENDOTHELIAL (R.E.) CELLS

The reticulo-endothelial cells or histiocytes may be divided for the convenience of description into two groups—*fixed* and *wandering* (Cf. chart, p. 104.)



I Fixed R.E. cells

(1) OF THE COMMON CONNECTIVE TISSUES ("TISSUE" HISTIOCYTES) and of the loose tissue of the serous membranes, e.g., omentum, pleura, etc. These are also sometimes referred to as "resting wandering cells" Their morphological characteristics are various. Some are round or spindle-shaped, others are squamous, while many have long mobile processes. They lie among the fibroblastic elements and often can be distinguished only with difficulty from the latter except by their special staining reactions. They may at any time as a result of some stimulus, particularly one of an inflammatory nature, become free and wander through the tissues. After the stimulus has been removed they may again come to rest.

(2) OF THE RETICULUM of the spleen, lymph glands, and bone marrow. These are large cells joined to one another by means of long branching processes. They lie among, and are attached to, the fibers of the reticular stroma. They too, given the necessary stimulus, may become detached and actively motile.

(3) FLAT ENDOTHELIAL-LIKE CELLS lining the blood sinuses of the spleen, bone marrow, adrenal cortex and pituitary. This group also includes those most interesting structures—the large flattened stellate cells in the blood sinuses of the liver (Kupffer cells). The latter possess many branching processes and project into the capillary lumen. In many instances they are almost free, being moored to the capillary wall by a delicate strand of protoplasm. At other times they may become detached and are carried away in the blood stream (fig 11 1).

(4) MICROGLIA of the central nervous system

II Wandering R.E. cells or free histiocytes

(1) OF THE SOLID TISSUES. From the foregoing it is seen that many of the fixed histiocytes may upon occasion become actively motile. Large cells of this type are found in the general connective tissues, in the omentum, in the splenic pulp and in the lymph glands,

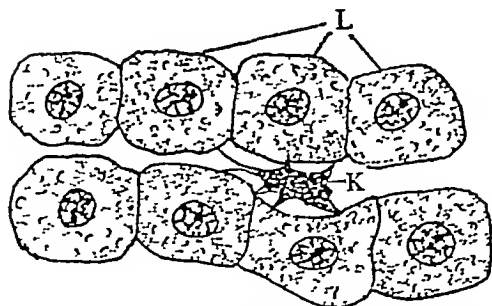


FIG. 11 1 Liver from an animal after the intravenous injection of India ink. K, Kupffer cells loaded with ink particles, L, liver cells

bone marrow, etc. (fig 11 2). The wandering cells may on the other hand come to rest and become fixed for the time. On account of the many different forms which these wandering macrophages may assume as a result of an inflammatory stimulus (p 107) Maximow has named them *polyblasts*.

(2) OF THE BLOOD (a) *Of extraneous origin*. Under certain circumstances as a result of some intense pathological stimulus (e.g., leukemia, bacterial endocarditis), the ordinary tissue macrophages already mentioned may be found in large numbers in the blood stream. They arise chiefly from the spleen and bone marrow and swarm into the venous system, but rarely reach the arterial side for, on account of their huge size (30μ), they are strained out by the capillaries of the lung. Large numbers may be obtained from the right heart but few if any from the left.

(b) *Monocytes* (see p 92). The origin of these normal elements of the blood has been the subject of considerable controversy. According to some, they arise solely from histiocytes, especially of the spleen and bone marrow, and are therefore classed with the reticulo-endothelial system. Lewis has shown that in hanging drop cultures macrophages may be transformed into monocytes, or vice versa. Maximow, on the contrary, believes that the monocytes arise from lymphocytes, stating that in certain situations, e.g., in the spleen, all transitional stages from lymphocyte to monocyte may be seen.

FUNCTIONS OF THE RETICULO-ENDOTHELIAL SYSTEM

The word "endothelial" is not quite appropriate. The cells lining the blood sinuses are not true endothelial cells. The ordinary endothelial cell of the blood vessels is a highly differentiated cell like the fibroblast and has no characters in common with the so-called endothelial cells of the sinuses of the bone marrow, spleen, etc. These are called *littoral* (shore) cells by Maximow. They are relics of the primitive endothelium.

From the account of the histiocytes which has been given it is quite evident that phagocytosis is one of their chief functions. In this they constitute one of the most important and powerful means by which the defense of the body is sustained. Their action, though similar to that of the neutrophilic leukocytes, is less mobile and more localized in character. They, with the aid of the lymphocytes, contribute toward the repair process which follows the acute phase of a tissue injury. The various types are for the most part readily transformable, one into another, the different forms which they assume being determined by local environmental conditions and the nature of the

stimulus. Any of the stationary cells may change into wandering histiocytes (macrophages) or mobile cells may become fixed and either retain in the sessile state their special phagocytic properties or lose these entirely and become converted into the ordinary connective tissue elements—fibroblasts—or into epithelioid cells. The fibroblasts, however, never undergo the reverse change and assume ameboid characters, once formed they remain fixed. According to Maximow, lymphocytes may give rise to macrophages. The latter, however, never give rise to lymphocytes.

In chronic inflammation or in the repair stage of an acute process, the histiocytes play an important rôle. Some (the so-called *dust cells* of the lung) are responsible for the removal of foreign particles which have been carried into the pulmonary alveoli by the inspired air. The ability of the Kupffer cells to take up ingested thorium dioxide (thorotrast) which is opaque to the X-rays is made use of to delineate the liver in the living subject. The spleen, placenta, ureter and kidney pelvis, the vessels of the extremities or of the brain and the cerebral ventricles, can also be outlined radiographically by means of this agent. The administration of thorotrast is not, however, free from danger. It is radioactive, the alpha ray activity of 25 cc of thorotrast being equivalent to that of 1 microgram of radium which is sufficient to produce pathological changes in susceptible persons. Thorium dioxide is eliminated from the body in insignificant amounts, almost all being permanently stored in the reticulo-endothelial cells of the bone marrow, spleen and lungs, as well as in the Kupffer cells, and even though no deleterious effects result from its radioactive properties it sets up a proliferation of connective tissue by acting as foreign material.

The epithelioid and giant cells of certain specific inflammatory processes, e.g., tubercle, which are derived from the histiocytes is but another instance of the latter's protean nature, and it is owing to these activities that the tissues are rendered so remarkably adaptable and plastic in their reactions to altered conditions. The omentum for example has long enjoyed a reputation as a protective structure, owing to its ability to form adhesions which serve to seal perforations of the gastrointestinal tract or to isolate infected regions within the abdominal cavity. The omental tissue is particularly richly supplied with both stationary and wandering histiocytes. It contains also, even in health, immense numbers of lymphocytes, which, in part, are derived from division of their own kind locally, and, in part, have come from the blood stream. Normally the presence of these various cells in such numbers gives an appearance closely resembling a defense reaction—the "physiological inflammation" of Rossle. Histiocytes and lymphocytes are continually being cast in showers

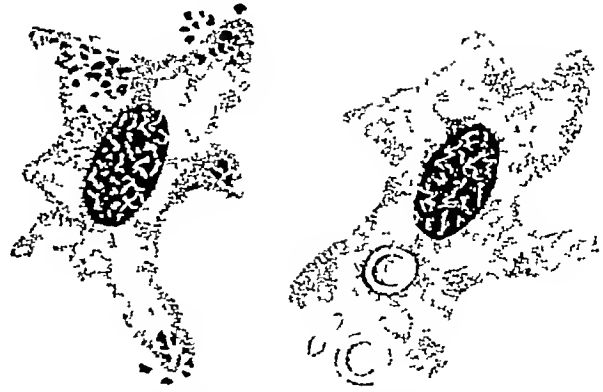


FIG 11.2 A, macrophage loaded with particles of India ink, B, macrophage showing processes, C, red cell (semi-diagrammatic)

into the peritoneal cavity at all times but to a much greater extent in irritative conditions. Monocytes have been studied in transparent chambers inserted into the tissue of the rabbit's ear. They have been observed to leave the circulation and to become motile and phagocytic, and otherwise indistinguishable from tissue histiocytes.

The formation of bile pigment (p. 537) and the final destruction of the blood cells (p. 73) in the spleen are other well-established functions of the histiocytes. The reticulo-endothelial system is also believed to be concerned with antibody formation.

In certain pathological conditions, the lipemia of diabetes, Gaucher's disease and Niemann-Pick's disease (p. 73) the reticulo-endothelial elements, especially of the spleen, are markedly increased in number and become loaded with lipid material.

The anti-reticular cytotoxic serum (A.C.S.) of Bogomoletz

The Soviet scientists, A. A. and H. O. Bogomoletz and their followers, believe that the reticulo-endothelial system and mesenchyme cells in general are of fundamental importance in the promotion and maintenance of bodily vigor and well-being, through their ability to raise tissue resistance in various pathological states and strengthen reparative processes. Directed by the theory suggested by Metchnikoff that, whereas large doses of a serum containing antibodies produced by the injection into animals of reticulo-endothelial cells are destructive to the corresponding cells in another animal, small doses of the antibodies are stimulating. These workers injected animals with extracts of organs of mesenchymal origin and rich in reticulo-endothelial elements, e.g., liver, spleen and bone marrow. They claim to have demonstrated a salutary effect of small doses of the tissue-specific antibodies produced in this way in such diverse conditions as scarlet fever, hypertension, carcinoma, erysipelas, rheumatism, tuberculosis and

certain mental states. The serum is also said to mitigate the degenerative changes due to age and to stimulate the healing of fractures. Though the latter effect has been confirmed by Straus and others, further confirmation of the beneficial effects of A.C.S. from other laboratories is required.

THE BLOOD-FORMING ORGANS OF POST-NATAL LIFE

Red cells, granulocytes and possibly platelets are formed in the bone marrow, and lymphocytes in the lymphoid tissue. The stroma or reticulum of these structures is the essential blood-forming tissue. The latter consists of (1) a fine network of cylindrical or ribbon-like fibers which can be stained by silver preparations and (2) large cells—*reticular cells*—which are fused together by branching processes to form a loose network, the latter is intimately associated with the network of fibers (1) which appear as reinforcing strands. The reticular cell (embryonic reticular cell, p. 104) is the nearest approach in the adult body to the primitive mesenchyme cell. The blood-forming capacity of the myeloid and lymphoid tissue depends upon the ability of these cells to form hemocytoblasts which in turn, according to the environment in which they are situated, are capable of developing into red cells, granulocytes or megakaryocytes on the one hand or lymphocytes on the other.

THE BONE MARROW

The red bone marrow is the hematopoietic tissue for erythrocytes and granulocytes in post-natal life. In the adult, this tissue is almost entirely confined to the flat bones, such as, the sternum, ribs, diploe of the skull and the bodies of the vertebrae. There is little or none, usually, in the long bones, such as, the tibia and the bones of the forearm, though small amounts may be found in the ends of the femur and humerus in young robust adults. The total volume of red bone marrow in the adult human body amounts to about 1400 cc. In infancy red marrow is present, not only in the ends of the long bones but fills the cavity of the shaft which in adult life contains only a yellow fatty material, this is not hematopoietic. But even in the adult the fatty marrow retains the essential reticular structure of hematopoietic tissue and is, therefore, capable under appropriate stimulation of blossoming into red marrow. So, under certain conditions, for example in pernicious anemia, and to a much less extent during residence at high

altitudes, or in any state associated with long-continued oxygen lack, the red marrow is increased in amount. It encroaches upon the medullary cavity of the shaft, replacing to a greater or less degree the fatty tissue.¹ In aplastic anemia on the other hand the fatty marrow increases at the expense of the red marrow.

In old age the volume of the red marrow is much reduced, most of the flat bones containing minimal amounts.

Drinker has shown that the circulation of the marrow is closed, i.e., the blood and the reticular tissue are separated by a complete membrane. This is contrary to the older view that the blood came into direct contact with the marrow cells through gaps in the capillary walls. The vascular bed of the marrow is a mesh-work of small blood sinuses (sinusoids) lined by flat endothelial-like cells similar in character to those lining the sinuses of the spleen (p. 71). Not all the sinusoids, however, are open at one time. It has been estimated by Doan that the marrow contains a great many more of these vessels than could possibly be accommodated within the resistant bony encasement, were they all in the dilated state. Some of the sinusoids are completely collapsed and impervious to the blood. Others are dilated but, owing to the vessels which lead to and from them being constricted, they are isolated from the general circulation. Regions of low oxygen tension are in this way provided. It has already been pointed out (p. 12) that a low oxygen tension serves as a stimulus to red cell formation.

The blood cells, if the observations of Maximow be accepted, are formed extravascularly and subsequently enter the blood stream.² Their pressure upon the endothelial wall causes its erosion or rupture and the cells invade en masse the lumen of the sinusoid where further growth and maturation of the cells follows. Finally, through the opening up at certain periods of the vessels leading from the hemopoietic region, the mature cells escape into the general circulation. The sudden showers of young cells that occur in pernicious

¹In some cases of rapidly developing anemia fat droplets may appear in the peripheral blood, and after some cases of hemorrhage the fatty acids of the blood increase. These facts suggest a dispersal of the medullary fat in order to make room for an extension of the blood-forming tissue—the red marrow.

²Sabin, Doan and others contend that only the granulocytes have an extravascular origin, the erythrocytes being derived, not from reticular cells but from the endothelial elements (littoral cells of Maximow) of the marrow capillaries.

anemia (blood crises) may be explained by a process of this nature. That the delivery of red cells into the circulation is normally intermittent is also suggested by the fact that the red cell count is not constant but shows a diurnal rhythm, the highest counts occurring in the morning hours. There may be a difference of 300,000 cells or so between the maximal and minimal daily levels of the red cell count. Normally no cells enter the general blood stream until maturation is nearly complete. Drinker, for instance, was unable to cause the passage of immature forms into the blood by prolonged perfusion of the marrow or by muscular exercise. The perfusion fluid evidently was unable to open a way into these isolated pools and wash out the immature forms. From what we know of the capillary circulation in other situations this result is not unexpected (ch 28).

Since the introduction of the technic of bone-marrow puncture and of culture methods our knowledge of the cellular structure of human marrow under normal and pathological conditions has been greatly advanced. The total number of nucleated cells per cubic millimeter in the bone marrow is around 75,000. Cells of the granulocyte series make up about 25 per cent and those of the erythrocyte series from 50 to 60 per cent. The remainder are megakaryocytes and unidentified cells. (See Plate 1)

THE MATURATION OF THE BLOOD CELLS

(a) Erythrocytes—Erythropoiesis

The red cells pass through several stages before they attain full maturity. In the examination of a simplified (hypoplastic) marrow, such as may be induced in the pigeon by underfeeding or in mammals after poisoning with benzol, most of the stages can be followed (cf fig 113 and frontispiece). The *reticular* cell, as we have seen, gives rise to the *hemocytoblast*. The latter divides into two daughter cells which stain deeply with basic dyes. These are called *basophilic erythroblasts* by Maximow, they contain no hemoglobin. The next stage which may be distinguished is, following Maximow's terminology, the *polychromatophil erythroblast*. The earliest cells of this stage are large, with a large, round and often vesicular nucleus, their cytoplasm is rich in basophilic material but also contains traces of hemoglobin. The hemoglobin concentration increases in amount as development advances, the more mature cells

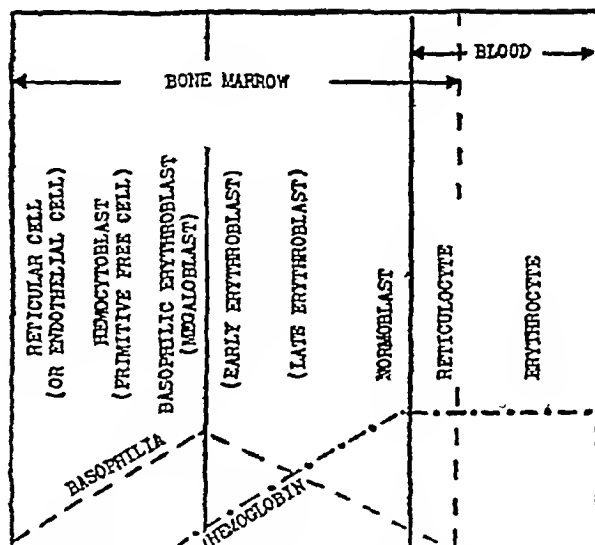


FIG 113 Diagram showing maturation of the red cells. The view of Sabin, Doan and associates is indicated by the parenthesized terms (see footnote, p 108). As mentioned in the text the early and late erythroblasts of these observers are termed polychromatophil erythroblasts by Maximow. This term has been omitted from the diagram for the sake of simplicity (modified from Sabin).

being well supplied with pigment. Hemoglobin is acidophilic, so the cells of the erythroblast stage, since they have retained a relatively large amount of basophilic substance, stain with both acid and basic dyes. This property of dual staining, known as *polychromasia*, diminishes again as the cells mature beyond the erythroblastic stage and gradually lose their basophilic material.

Several authors (Sabin, Doan and associates) distinguish three types of cells during this stage of development—the *megaloblast* and the *early* and *late erythroblasts*. The megaloblast of these authors corresponds to the youngest of the polychromatophil erythroblasts mentioned above, i.e., a large nucleated cell which, owing to its relatively large proportion of basophilic material, shows only a very slight tendency to polychromasia. Maximow avoids the term megaloblast, advising that it be reserved for the large cell of pernicious anemia to which it was originally applied. He as well as other hematologists claim that the latter cell, though resembling the primitive cell of normal marrow, is not identical with it (see p 84).

The older erythroblasts give rise to *normoblasts*. The normoblast, as the name itself implies, resembles the mature erythrocyte in size and hemoglobin content but still retains its nucleus which, however, shows condensation of its chromatin material (*pyknosis*) and stains more deeply. In the final stage of the maturation process, the

nucleus is extruded from the cell and the now nearly mature erythrocyte is discharged from the marrow into the general circulation. It betrays its youth only by a fine basophilic reticulation of its cytoplasm and is therefore called the *reticulocyte* (p 15). It is probably not until the reticulocytes have reached a certain concentration in the marrow that they become discharged into the general blood stream.

In healthy marrow the multiplication of the red cells which occurs to replace those lost from the circulation through wear and tear is effected almost entirely by the division and redivision of later forms, i.e., of normoblasts and older erythroblasts (homoplastic development) and to a negligible extent through the multiplication and subsequent maturation of the more primitive forms. The youngest erythroblasts (basophilic and polychromatophilic erythroblasts), for example, in normal marrow are absent or amount to no more than 0.01 to 0.04 per cent of the total nucleated cells of the erythrocyte series, according to Sabin, the older erythroblasts make up about 30 per cent and the normoblasts 70 per cent or so. In pernicious anemia, on the other hand, large numbers of very primitive cells including the characteristic megaloblast are present. In this disease erythropoiesis is abnormal in that the cells produced are not only of a more primitive type but are different from ones ever seen at any stage of normal erythropoiesis. It is not then merely a matter of reversion to a primitive but otherwise normal mode of red-cell development. The benefits of liver therapy apparently depend upon the power of the hepatic principle to restore the normal mode of red cell development.

(b) Granulocytes—granulopoiesis

The earliest stage in the differentiation of the granulocytes from the primitive reticular cell of the marrow is that of the so-called *myeloblast*. It is generally agreed that the origin of the granulocytes is extravascular and that they pass into the marrow vessels only after they acquire motility. Maximow considers the myeloblast to be nothing more or less than the polyvalent hemocytoblast or stem cell, identical with the large lymphocyte, the minute differences in structure between it and the latter being due, it is thought, to the environment in which it is placed, i.e., the bone marrow. Here it gives rise to the three types of granulocyte. The myeloblast is given added interest since a cell

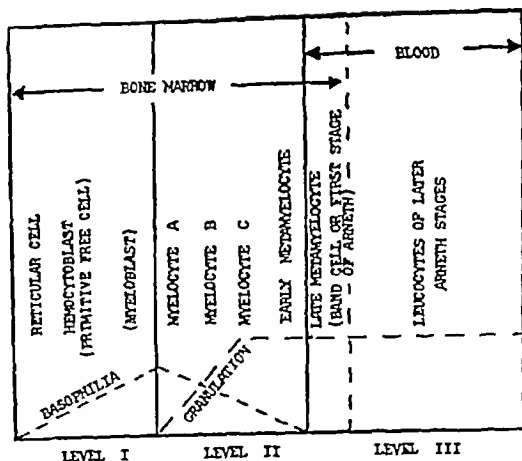


FIG 11.4 Diagram showing maturation of the granulocytes. Parenthesized terms are those of Sabin, Doan and associates (modified from Sabin).

similar to, but not identical with it, is predominant in the blood and marrow of certain forms of *leukemia* (p 99) when it may constitute 90 per cent of the white cells. It forms only a very small percentage of the cells of normal marrow.

After the stage of the myeloblast a few specific granules begin to appear in the cytoplasm which at the same time becomes less basophilic in character. Also present are a few azurophil granules. In the subsequent history of the cell up to its discharge into the circulation four stages, based upon the development of specific granulation and the reduction in basophilic material, are recognized. The cells of the first three of these stages are termed *myelocytes A, B and C*, respectively. The cells of the fourth stage are called *metamyelocytes* (fig 11.5). In the youngest stage (A) the granules, as mentioned above, are very few³ and are in greater numbers in myelocyte B and in maximum quantity in myelocyte C. The nucleus in these stages is oval or spherical and shows no attempt at lobulation. The metamyelocyte is characterized by slight indentation of the nucleus and the first evidence of typical ameboid movement. Finally the nucleus becomes deeply indented or constricted at one or more points and is discharged into the circulation as a young leukocyte (neutrophil, eosinophil or basophil). This is the first stage of Arneith mentioned on page 93. It is also

³ This cell, myelocyte A, resembles closely its progenitor the myeloblast and represents a transition stage between the latter cell and the older myelocyte. It is often referred to as the promyelocyte. Some use this latter term to include also myelocyte B.

referred to by Schilling as the older metamyelocyte or the band cell. As shown in figure 114 three levels of development are distinguished. Level I extends from the reticular cell to, and includes, the myeloblast. Level II embraces the myelocyte and the metamyelocyte, cells at this level multiply actively and have the power of growth. Level III includes those forms present in normal blood (Arneth stages), the cells have lost the power to divide and grow.

As in the case of erythropoiesis the earliest stages of white cell development—myeloblastic and early, myelocytic—are little in evidence in normal bone marrow. The leukocytes are supplied to the blood through the division and redi-
vision of the later forms of white cell elements—i.e., the myelocytes. A census taken of the colorless elements of the marrow in normal rabbits gives an average of about 1 per cent myeloblasts, 90 per cent myelocytes and metamyelocytes, of which the great majority are neutrophilic, and about 4 per cent young leukocytes. Younger myelocytes make up the balance. The number of leukocytic elements is three times greater than the number of red cell elements—erythroblasts and normoblasts. The ratio of granulocytes to red cells in blood, on the other hand, is 1 to 600. The difference in the cell ratios in marrow and in blood is ascribed to the greater mortality of the leukocytes and the consequent necessity for a greater hemopoietic activity in order to replace them.

In myeloid leukemia the process of granulopoiesis is pushed back to a more embryonic type and immature forms appear in the blood. The marrow picture indicates a great activity of the early myelocytic and myeloblastic stages. The marrow of the fetus and to a less degree that of the new-born show also a relatively large proportion of immature cells.

(c) *Lymphoid tissue and the formation of lymphocytes*

The central areas of the follicles in lymphoid tissue (lymph glands, spleen, etc.) stain more lightly with the ordinary stains than the peripheral zones. These lighter areas, about 1 mm in diameter and pierced near the center by a small arteriole, are generally known as the "germ centers." In the embryo and in new-born animals the earliest stage in the development of the lymphocytes is represented in lymphoid tissue as well as in the bone marrow by a large lymphocyte or hemocytoblast. It is analogous and practically identical

in structure with the myeloblast. It is called the *lymphoblast*. In postnatal life, lymphocytes arise only from lymphoid tissue. The ancestral cell or hemocytoblast stems from the undifferentiated primitive reticular cell of the lymphoid stroma similar to that in the marrow from which the myeloblast and, according to Maximow, the erythroblast arise but, as in the case of the myeloid cells, the lymphocytes normally arise from the division of cells of later stages.

The total mass of lymphoid tissue amounts to between 1 and 3 per cent of the body weight, and its output of cells is relatively enormous. Yoffey estimated the daily production at 35,000,000,000, the tremendous wastage has already been mentioned (ch. 8).

An interesting experiment of Maximow's supports the view that the large lymphocyte or hemocytoblast is the common precursor of the granulocyte and lymphocyte and that the course of development which the stem cell shall follow is determined by its immediate environment. Lymphoid tissue was cultured in an environment prepared to simulate that of the bone marrow by the use of a preparation of blood plasma and marrow extract. Proliferation of the large lymphocytes and their differentiation into myelocytes were clearly observed.

DISTINGUISHING FEATURES OF THE DIFFERENT STAGES OF THE LEUKOCYTE SERIES

OXIDASE GRANULES

The mature granulocytes, the myelocytes and myeloblasts (except the very youngest) and the primitive monocytes, contain fine, reddish-brown granules (*oxidase granules*) which become evident when the cells are stained in the following way (Washburn technique). The cells are first treated with a solution consisting of benzidine, 0.3 gm. in 99 cc. of alcohol and 1 ccm. of a saturated solution of sodium nitroprusside. Ten drops of this solution are applied to a dried blood film and allowed to stand for 4½ minutes. Five drops of a solution of hydrogen peroxide in 25 ccm. of water are then added. After standing for another 4½ minutes the film is washed and dried and stained with Lieshman's stain. Cells of the lymphoid series do not contain oxidase granules.

MOTILITY AND PHAGOCYTOSIS

When stained supravitaly the primitive monocyte is phagocytic and actively motile, whereas myeloblasts and lymphoblasts are usually stated to be non-motile.

In hanging drop preparations, however, Rich states that both of the last mentioned cells also show definite motility and that the mode of locomotion of each type is characteristic. The myeloblasts and the younger myelocytes exhibit a worm like motion, whereas the lymphoblast protrudes a long tongue of cytoplasm in a direction opposite to that in which it is moving, it thus assumes a shape resembling that of a hand mirror or a frying pan.

NUCLEAR STRUCTURE

The nucleus of the myeloblast shows a delicate reticular structure and no definite nuclear membrane, the nucleoli are indefinitely marked off from the rest of the nuclear material. The nucleus of the lymphocyte on the contrary is well defined, as are also the nucleoli, and the reticulum is coarser.

These distinguishing features of the various cells types are employed in diagnosing the different forms of leukemia.

(d) *The platelets*

These are generally stated to arise from giant cells of the marrow (40μ or more in diameter) known as megakaryocytes. The latter have an irregular, ring shaped nucleus, and are capable of amoeboid movements. It was first suggested by Wright that fragments of their protoplasm became detached to form the blood platelets. Normally, the marrow contains only a few of these cells, but it has been stated that an increase or decrease in

their number is followed by corresponding changes in the number of circulating platelets.⁴

(e) *The origin of monocytes (see p 106)*

⁴Wright's view of the origin of the platelets from megakaryocytes, though very widely held, has not gone unquestioned. Various alternate theories have been advanced with respect to the origin of the platelets. Some believe that they are formed from the cytoplasm of leukocytes, and others that they are simply precipitates from the plasma. Howell and Donahue concluded from their experiments that the platelets are derived from megakaryocytes in the lungs. They based their view on the finding that both in cats and in man the platelet count was somewhat higher in arterial than in venous blood. The ratio of platelets to erythrocytes in arterial blood was 1:21.4, as compared with 1:23.8 in the blood of the corresponding vein, which implies that platelets are destroyed or removed from circulation in the capillaries and replenished in the lungs. Further evidence was derived from the observation that a smear of lung tissue showed large numbers of giant cells, whereas only a few were found in a similar specimen of bone marrow. It is probable, however, that the lung megakaryocytes do not originate in the lungs but have their source in the bone marrow and, owing to their large size, are merely trapped in the pulmonary capillaries. Several observers have expressed the belief that they are fragments of degenerated red cells. In support of this view Watson reports the observation that phenylhydrazine which damages the erythrocytes causes a sharp rise in the platelet count and a reduction in the number of red cells. This observer also found that, when erythrocytes were suspended in a counting chamber, degeneration of the red cells occurred while bodies indistinguishable from platelets appeared. Furthermore, the platelet count, as shown by Bedson, is increased by splenectomy, or by "blocking" the reticulo-endothelial macrophages (which, as we have seen, engulf red cell fragments) by injections of India ink. The term "blocking" refers to the overloading of the reticulo-endothelial cells with the ink particles and the consequent suppression of their phagocytic properties.

CHAPTER 12

THE COAGULATION (CLOTTING) OF BLOOD

GENERAL DESCRIPTION OF CLOT FORMATION

If blood is collected into a test tube it will be found at the end of 5 or 6 minutes that it has lost its fluidity and has set into a jelly. The tube may be inverted, but the blood, which is now said to have *clotted* or *coagulated*, does not escape. If it were possible to magnify this clot many times and to look within it, one would see a mesh of very delicate fibrils, among which were entangled, as in a net, the red and white cells and many fragmented platelets. The fibrils can be readily revealed when a thin section of the clot is examined under the high power of the microscope. They are composed of *fibrin* formed by the conversion of the fibrinogen of the plasma from a soluble (hydrosol) into an insoluble form (hydrogel). The fibrin forms ultra-microscopic crystal-like needles which, as they are deposited, create tenuous, interlacing filaments within the structure of the protein. The process by which this change of fibrinogen from a liquid to a more or less rigid structure (fibrin) is effected has been compared by Mommaertz to the manner in which a matted felt-like mass is formed from elongated objects—as reeds floating in water become tangled and coherent. It will be recalled that fibrinogen itself is composed of long fiber-like molecules. The immobilization of the molecules in this manner is called *coacervation*¹ by de Jong.

If the clot is permitted to stand for an hour or so, it will be found to have shrunk, and in shrinking to have expressed from its interstices a clear, faintly straw-colored fluid. This is the *serum*. The latter remains perfectly fluid and is quite incapable of clotting. The shrinking and condensation of the clot is due to the gradual shortening of the fibrin threads which enmesh the corpuscles. In this retraction of the clot or *syneresis*, as it is termed, the platelets play an essential rôle. Though numbers of platelets undergo disintegration when the blood is shed (see below), others become attached here and there in groups or knot-like clumps to

the fibrin threads, and in some unexplained way cause bending and shortening of the latter. Clots formed in blood deficient in platelets are soft and friable, and do not retract in the normal way. It used to be thought that the intact platelets which were seen in sections of blood clots served as nuclei from which the fibrinogen to fibrin conversion started, but it has been demonstrated by Tocantins that the platelets take up their positions after the fibrin has formed. The full retraction of the clot and the separation of the serum ordinarily takes a considerable time, but separation can be brought about within a few minutes by rapid centrifuging.²

If the blood be centrifuged as soon as it is shed and the cells in this way separated from the plasma, a clot forms in the latter, due as before to the formation of fibrin threads. After a time the colorless clot shrinks and, as in the case of whole blood, expresses the transparent serum. The clotting process is therefore essentially a phenomenon of the plasma. The straw-colored layer—the so-called “buffy coat”—which forms on the surface of blood which clots slowly, and so allows a certain degree of sedimentation of the cellular elements to occur, is clotted plasma.

THE CLOTTING MECHANISM

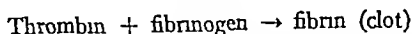
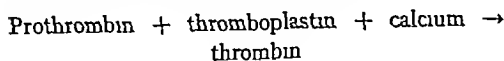
In the very intricate mechanism underlying the coagulation of blood the following substances take part (1) *Prothrombin*, (2) *thrombin*, (3) *thromboplastin* (or *thrombokinas*), (4) *ionized calcium* and (5) *fibrinogen*. By the action of thromboplastin, prothrombin, which is inactive, is converted in the presence of calcium ions into the active thrombin. Thrombin then acts upon the

² There is a direct relationship between the platelet count and the rapidity and degree of clot retraction and the resulting firmness of the clot. The effect of the platelets upon clot retraction is exerted only when they disintegrate. The concentration of thrombin, which induces lysis of these elements, therefore influences the retraction of the clot. The concentration in the blood of the other formed elements affects clot retraction in the opposite way. In anemia, for example, clot retraction is more pronounced than normally, whereas, in polycythemia the clot is soft and friable. Increase in the fibrinogen concentration acts in a manner similar to that caused by a high platelet count, hastening the retraction of the clot.

¹ De Jong defines this as the process occurring in a condensed colloidal system, in which particles with positive and negative electric charges are mutually attracted. According to the view of Mommaertz, a pro fibrin is first produced and then, through electrostatic attraction of its particles, a mesh-like interlacement occurs to produce fibrin.

soluble fibrinogen of the plasma to convert it into the insoluble fibrin spoken of above as forming threads in which the solid elements of the blood are enmeshed. Thus the clot is formed.

The clotting mechanism in the simplest possible terms is shown in the following scheme



These four primary factors have long been recognized as being fundamentally involved in the coagulation of the blood.

Prothrombin and ionized calcium are present in the circulation, but thromboplastin, though contained in the solid tissues, is virtually absent from the blood.

Normally, the blood remains fluid in the vessels, not, as was thought at one time, because it is in motion, but most probably for the reasons that thromboplastin is present in the circulation in only minute amounts, or in an inactive form, and that the platelets remain intact. Thrombin, the coagulating enzyme, is, therefore, not formed, moreover, an antithrombin is present in normal blood, which serves as a safeguard against any thrombin which might arise in sufficient amount to cause intravascular clotting. This material, the so-called normal antithrombin of the plasma, is present in low concentration in mammalian blood but in much larger quantities in the blood of birds. According to Quick, it is closely associated with the albumin fraction of the plasma.

No field in physiology is so complex and has been so confused by contradictory results, differences of interpretation and diversity of opinion than that of blood coagulation. Each fact discovered, instead of simplifying the subject, usually brings new problems in its train. The relatively simple mechanism expressed above must now be elaborated in the light of modern research.

The clot accelerating factor in plasma (Ac-globulin) The work of several laboratories had indicated the existence in blood of a factor which accelerated the conversion of prothrombin to thrombin in the presence of ionized calcium and thromboplastin. In 1943 Quick concluded from his experiments that prothrombin was not a single substance but consisted of two components, A and B. Component A disappeared from stored plasma—component B, which could be adsorbed by aluminum hydroxide ($\text{Al}(\text{OH})_3$), was diminished

by dicoumarol (p. 118) and in hypoprothrombinaemia caused in other ways. Component A was later called the *labile factor* by Quick,³ and is considered to be identical with a factor subsequently reported by Owren, as well as by Fantl and Nance, and by Seegers and his associates. Fantl and Nance described this factor in plasma. Owren named it factor V. This substance is also called *plasma accelerator globulin* (Ac-globulin) by Seegers and his colleagues, who believe it to be an inert protein or pro-enzyme. According to these latter workers, plasma Ac globulin is inactive but is converted to an active form, *serum Ac-globulin*, by small amounts of thrombin.

The role of the platelets Contrary to previous beliefs Quick has shown that the platelets contain little or no thromboplastin, yet the platelets are responsible, indirectly, for the production of thromboplastin. The latter substance, it is postulated, is present in the plasma in an inactive form which has been named *thromboplastinogen*. It is converted to active thromboplastin by a factor liberated from lysed platelets and known as *thromboplastinogenase*. The disintegration of platelets is a prerequisite for the production of thromboplastin and, therefore, for the clotting of blood. Thrombin in minute amounts causes the lysis of platelets. Several anticoagulant agencies prevent coagulation by retarding platelet dissolution.

Prothrombin is also present in blood in the form of a precursor known as *prothrombinogen*, which is converted to prothrombin by contact with glass or any rough surface. The mechanism of this conversion is obscure.

A general summary of the clotting mechanism can now be given.

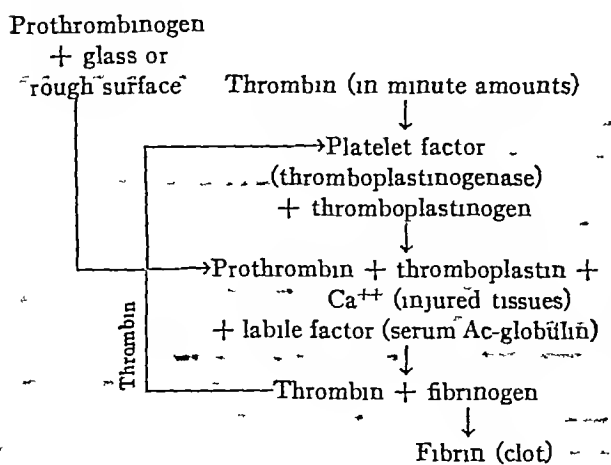
The coagulation of blood involves a series of autocatalytic reactions—a sort of chain-reaction which once started continues to completion. Though every step in the complex process has not been firmly established beyond all possibility that at some time new discoveries will not require modifications to be made, the sequence of events is probably as follows. When blood is shed and comes into contact with glass or a rough surface, a certain amount of prothrombinogen is at once converted to prothrombin and a few platelets are broken up. A small amount of thrombin is therefore

³ Quick states that the labile factor (Ac globulin) is essential for the *prothrombin + thromboplastin + Ca^{++} → thrombin* reaction, and is not merely an accelerator.

produced which brings about a greater lysis of platelets. The platelet factor (thromboplastinogenase) thus liberated converts thromboplastinogen to thromboplastin and if the blood comes into contact with damaged tissues more thromboplastin is made available. This substance in the presence of Ca^{++} and the labile factor reacting with prothrombin produces thrombin which acts upon the soluble fibrinogen to form insoluble fibrin. Part of the thrombin causes a further disintegration of platelets and the production of more prothrombin from its precursor to sustain the clotting mechanism.

Quick doubts whether the clot has much immediate mechanical effect in arresting the bleeding, that is, of acting as a plug to close the wound in the vessel. Factors of greater importance, as stated on p. 28, in bringing about hemostasis are, reflex contraction of the vascular wall, the liberation of vasoconstrictor and hypotensive substances from the platelets, and the stickiness of the endothelial lining of the vessel, which by "glueing" together the apposed surfaces helps to seal the cut or torn vascular wall. Unless the blood flows freely to the exterior, an additional factor is the pressure of the blood extravasated into the tissues upon the bleeding vessel. The fibrin threads of the clot, apart from any effect they may have in acting as a plug, serve to provide a rough surface for the production of prothrombin from its precursor, to adsorb fibrin, and thereby to prevent the intravascular extension of the clotting process, and to trap clumps of agglutinated platelets which are thus held at the site of the bleeding to exert a persistent hemostatic action.

The several factors in the clotting mechanism are shown in the following scheme:



Heparin The action of certain impure extracts of liver in inhibiting the coagulation of blood was discovered by McLean, a pupil of Howell, in 1916. Later, a series of studies was carried out by Howell and Holt upon extracts of liver from which they obtained a powerful anticoagulant. They gave it the appropriate name of *heparin*. Howell and Holt obtained their material from dog's liver, but it was later prepared in larger quantities by Charles and Scott from beef liver. Lung, muscle and intestinal wall also contain it in relatively large amounts. Smaller quantities are contained in spleen, heart and thymus. It is present in negligible quantities, if at all, in normal blood. Heparin is active both *in vivo* and *in vitro*, when it is injected into the living animal the blood remains incoagulable for several hours. A unit of heparin is defined as the quantity of material which will prevent for 24 hours the clotting of 1 cc of cats' blood when kept in the cold. A very potent crystalline preparation has been obtained from beef liver by Scott and Charles. This contains nearly 100 units per milligram.

The anticoagulant action of heparin is complex and depends upon at least three effects which it exerts upon the clotting mechanism. 1. It is a powerful *antithrombin*, that is, it inhibits the action of thrombin in converting fibrinogen to fibrin. Quick found that for this action a co-factor associated with the albumin fraction of the plasma is necessary, it has been named albumin X. 2. Heparin is also an *antiprothrombin*, i.e., it prevents the activation of prothrombin to thrombin. This property has been questioned until recently, but the experiments of Ferguson and of Quick have firmly established this action of heparin. 3. Through its effects (1) and (2) in inhibiting the formation and action of thrombin, heparin reduces or prevents the agglutination and lysis of platelets and, as a consequence, the production (through thromboplastinogenase) of plasma thromboplastinogen.

Jorpes found that, like certain other anticoagulants, such as azo dyes and germanin, the heparin molecule contains sulphuric acid groups, in virtue of which it acts as a relatively strong acid. It belongs to the same chemical group as that to which mucic acid and chondroitin sulphuric acids belong. It is a complex carbohydrate containing esters of mucic acid-sulphuric acid and yields glucuronic acid, glucosamine and sulphuric acid groups upon hydrolysis. Anticoagulants have been prepared by Chargaff and his associates by the introduction of sulphuric acid groups into such polysaccharides as cellulose.

lose, glycogen and starch and into the cerebrosides, cerebrin and kersin.

Since heparin is absent from the circulation, or present only in minute amounts, it cannot be responsible for maintaining the fluidity of the blood in the living body. The most probable reasons that the blood does not clot intravascularly have been stated on page 114.

From the evidence procured by Wilander, heparin appears to be a product of the "mast cells" of Ehrlich (i.e., mobile cells in the tissues containing fine basophil granules) which may be seen in clusters around the minute vessels of those tissues which give a high yield of heparin. Mellanby suggests that the significance of heparin in tissues is that it serves as a *local* anticoagulant, preventing the clotting of blood circulating in the small vessels. The fine granules to be seen in these cells show the metachromatic staining reaction, turning violet when treated with toluidine blue or azure A. Heparin itself gives this reaction and the granules are therefore believed to be the anticoagulant in the form of minute droplets.

Chargaff and Olson found that protamine (salmine) annuls the action of heparin both *in vitro* and *in vivo*. Protamine, a simple protein with basic properties (see also p. 623) combines with heparin, the resulting compound being quite free from anticoagulant action. Thus it is possible to determine the quantity of heparin in a sample of blood by ascertaining the quantity of protamine required to give the shortest coagulation time. By this means the great lengthening of the coagulation time in anaphylaxis (up to 2 days or more) and peptone shock has been shown by Waters, Markowitz and Jaques to be due to the high concentration of heparin in the circulation (*Heparinemia*, p. 118 and p. 122). Heparin is most effective when administered intravenously, much larger doses being required when it is given by subcutaneous injections. Besides its action in prolonging the coagulation time, heparin, as shown by Murray and associates, hinders the agglutination and deposition of platelets and thus discourages thrombus formation. Heparin has been recognized for a number of years as a valuable anticoagulant for use in physiological experiments, but it is only comparatively recently that purified preparations suitable for clinical use have become available. Its most important field of usefulness is in blood transfusions, in operations upon the blood vessels and to check the extension of coronary, and certain

other types of thrombosis. Its employment as a preventive against postoperative and other thrombotic states which carry the threat of fatal pulmonary embolism (p. 124) has been advocated by several investigators (Howell, Mason, Hedenius, and Murray and associates). Widstrom and Wilander have suggested its use to prevent the formation of fibrinous adhesions in pleurisy, in heparinized rabbits the inflammatory exudate caused by the injection of iodine into the pleural cavity is incoagulable.

THE PROPERTIES OF PROTHROMBIN AND THROMBIN AND OF THROMBOPLASTIN

Prothrombin and thrombin, according to Seegers, are carbohydrate-containing proteins. This investigator has prepared from beef plasma highly potent preparations of these agents which contain, respectively, 300 units of prothrombin activity and 600 units of thrombin activity per milligram. At a pH of 7.0 prothrombin and thrombin are highly soluble in water or physiological saline, it being possible to prepare a 60 per cent solution of either material. Prothrombin is insoluble at a pH of 4.9 and thrombin at about 4.3. Both substances are completely and permanently inactivated by acid (pH 3.5) and by alkali (pH 11.0). They are destroyed by heating to 60°C for 30 minutes. The concentration of prothrombin in blood is around 25 mg per 100 cc.

Thrombin is an enzyme, but the conversion of prothrombin to thrombin by thromboplastin and ionized calcium, as well as the action of the labile factor, involve straight stoichiometric reactions.

A linear relationship (inverse law) can be shown between the relative concentration of thrombin in the plasma and the reciprocal of the coagulation time, i.e., the higher the thrombin concentration, the shorter the time required for a clot to appear (see fig. 12.1).

The origin of prothrombin. Prothrombin is formed in the liver and thus appears to be its principal source, though the experiments of Drinker point to its being produced to some extent as well by the bone marrow. This observer obtained a fluid rich in prothrombin by perfusion of the bone marrow. The production of prothrombin is governed by vitamin K (ch. 55).

Thromboplastin is present in all tissues, lung and brain being especially rich sources. Though it is the general custom to speak of thromboplastin as a single principle, there are actually at least

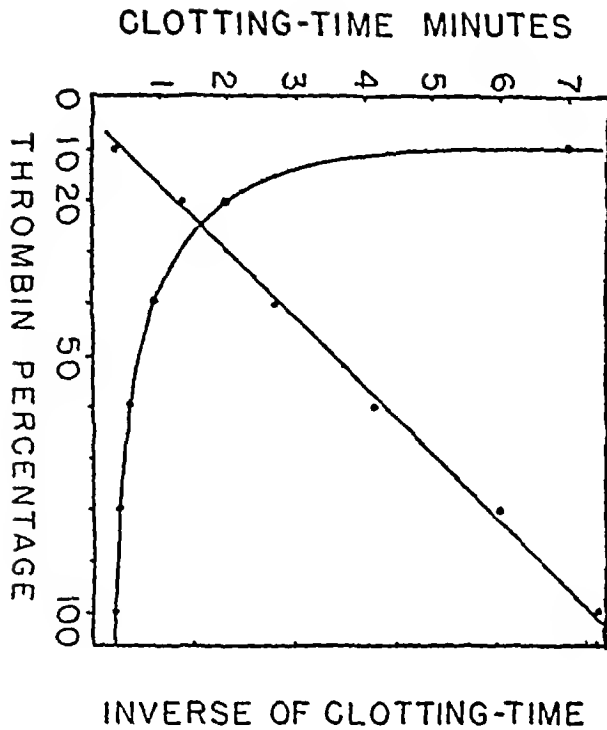


FIG 12 1 Relating clotting time in minutes for pure fibrinogen (ordinate) to varying concentration of thrombin (abscissa) (After Ferguson)

two classes of thromboplastic substances. To one of these classes belong the *phospholipids*, e.g., *cephalin*, first shown by Howell to have this action, and *lecithin*. They are fat soluble and present in relatively large amounts in brain and nervous tissues, though also contained in tissues generally. The other class are water soluble and chemically are *lipoproteins*. They are in especially high concentration in pulmonary tissue. The thromboplastin formed in plasma belongs to this latter class. The phospholipids are much slower in their action than the lipoproteins.

The means used to prevent or retard coagulation—anticoagulants

COLD Since the clotting mechanism consists of a series of chemical and enzymatic reactions, it is to be expected that coagulation will be retarded by lowering the temperature. Keeping blood at a temperature of from 5° to 10°C postpones coagulation but does not absolutely prevent its occurrence. Cold, e.g., ice, etc., applied to the surface of the body as a means of arresting hemorrhage has, however, no effect in retarding the coagulation process. Actually a hemostatic action is exerted in such instances as a result of the vasoconstriction reflexly induced.

(2) AVOIDANCE OF CONTACT OF THE BLOOD WITH FOREIGN MATERIALS OR INJURED TISSUES. Drawing the

blood directly from an artery or vein (in order to prevent contact with the tissues—a source of thromboplastin), and collecting it into a receptacle whose walls have been made smooth and *unwettable* by a coating of paraffin, or preferably of *silicone*, will slow greatly the coagulation process. Silicone is the most effective substance known for this purpose, it acts like the vascular endothelium in inhibiting the activation of prothrombinogen (to prothrombin) and tends to preserve the platelets (Jacques and associates). Plasma, after collection in silicone lined apparatus and centrifuged at high speed in order to remove the platelets as completely as possible, will remain fluid for 3 or 4 days even at room temperature. Avian blood collected in this way will remain unclotted almost indefinitely. The importance of a smooth surface in maintaining the fluidity of blood is also shown clearly in the classical experiment of the "living test tube." If a section of the jugular vein of an animal, e.g., the horse, be isolated between ligatures and carefully removed with its contained blood, this will not clot for a long period, not until changes occur in the lining of the vessel wall and the necessary thromboplastin is thereby provided.

On the other hand, if the blood flows over a rough surface, coagulation is hastened. The blood coming from a ragged wound which involves a greater destruction of tissue, is also likely to clot more quickly than if it issues from a blood vessel that has been cleanly incised, as by a scalpel or a razor. The contact of a sponge or a powdered substance to the wound by increasing the extent of surface exposed also hastens the formation of the clot.

(3) **DE-CALCIFICATION** The addition of oxalate (sodium, potassium or ammonium) or of a fluoride to blood to the extent of 0.1 per cent or more, completely destroys its power to clot spontaneously. In the former instance the calcium is precipitated as calcium oxalate. Oxalated blood recovers its ability to clot if shaken with chloroform, but the fibrin which forms redissolves after the blood has stood for a time. Fluoride does not precipitate the calcium but forms a weakly dissociated calcium compound. Sodium citrate also prevents coagulation. In this instance a double salt—calcium sodium citrate—is formed which again is only slightly dissociated. If calcium in the ionic form, e.g., the chloride, be added to the oxalated, fluoride or citrated plasma (recalcification), the power of the blood to clot is regained. Also the addition of thrombin or of thromboplastin to the decalcified blood causes it to clot.

(4) **NEUTRAL SALTS** Magnesium sulphate solution in a strength of 27 per cent mixed in the proportion of 1 to 4 of blood postpones coagulation for some time, but does not prevent it indefinitely. Sodium sulphate in half saturated solution added to an equal quantity of blood has a similar effect, as has also a 10 per cent solution of sodium chloride in the same proportion. The manner in which these substances act is not clear. They do not

decalcify the blood. That the activity of some or other elements necessary for the clotting process is merely suspended and that none is destroyed, is shown by the fact that mere dilution of the "salted" blood is followed by clotting. Some think that these substances may act by preserving the platelets and corpuscles from disintegration and so hindering the liberation of thromboplastin. Zinc sulphate in 0.5 per cent solution prevents coagulation by precipitation of the fibrinogen. Sodium thiosulphate and germanin are anticoagulant, this action depending apparently upon their sulphur content.

(5) AZO-DYES, such as Chicago blue 6 B, Trypan red, Trypan blue, Chlorazol fast pink, B.K.S. etc., are powerfully anticoagulant. Chicago blue, the anti-coagulant property of which was discovered by Rous and his associates in 1930, is more potent either *in vivo* or *in vitro* than the earlier preparations of heparin. One milligram of purified Chicago blue will prevent for 60 hours the coagulation of 1 cc. of rabbit's blood kept on ice.

(6) CERTAIN SUBSTANCES OF A BIOLOGICAL NATURE

(a) *Hirudin*, a substance secreted by the buccal glands of the leech and extracted commercially, has a very powerful anticoagulant effect. It is a typical anti-thrombin, i.e., it inactivates thrombin. Hirudin has been used extensively in the past to prevent coagulation during physiological experiments. (b) *Snake venoms*. The poisons of some snakes, particularly of the cobra, are powerfully anticoagulant. One hundredth part of a milligram per kg. of body weight will, according to Howell, entirely prevent coagulation. Its anticoagulant effect is thought to be due to some chemical change which is produced in the thromboplastin. The anticoagulant effect of cobra venom is probably intimately associated with its hemolytic action (p. 65). Some other snake venoms, however, have just the reverse effect and cause intravascular clotting. (c) *Heparin* (see p. 115).

(d) *Peptone solution* when injected prevents coagulation. According to Howell it acts by stimulating the liberation of heparin from the liver. Confirmatory evidence for this explanation has been furnished by others (Quick, and Waters and associates). Peptones also have an anticoagulant action when added to shed blood. The anticoagulant effect under these circumstances is ascribed by Mills and by Pickering to the increased resistance of the platelets to disintegration induced by the peptone solution. (e) In *anaphylactic shock* the blood remains incoagulable indefinitely (up to 2 days at least). Waters, Markowitz and Jaques have shown that this is due to the liberation of heparin from the liver. Anaphylaxis produced in hepatectomized dogs is not accompanied by a prolonged coagulation time, (f) *Cysteine* (see p. 546). (g) *Dicoumarol* 3,3'-methylene bis(4-hydroxycoumarin), present in spoiled sweet clover hay, see p. 120. Schofield in Canada and Roderick in N. Dakota showed that a hemorrhagic disease of farm animals could be cured by replacing spoiled sweet clover in the diet with fresh fodder. The latter observer also

showed that the reduced coagulability of the blood was due to diminution in prothrombin. Link demonstrated that the toxic principle was the substance now generally known as dicoumarol, a derivative of coumarin, ($C_9H_6O_2$) the odorous principle in new mown clover hay. Dicoumarol does not act *in vitro* as an anticoagulant nor in the body except after it has been partially metabolized. Link and his colleagues have procured strong evidence that in its metabolism salicylic acid is produced and is the active agent. When dicoumarol is administered intravenously, only traces can be detected in the circulation after from 20 to 24 hours, yet it is not excreted in the urine or feces. Most of the drug has disappeared before any appreciable reduction in the concentration of prothrombin in the blood has occurred. *In vitro* experiments have demonstrated the conversion of dicoumarol quantitatively to salicylic acid. It has long been known that the administration of salicylates, e.g., acetylsalicylic acid (aspirin) and sodium salicylate, is often followed by spontaneous bleeding, which has been shown to be associated with hyperprothrombinemia. Salicylates also cause a diminution in the number of platelets. The anticoagulant action of salicylates is antagonized by vitamin K.

Defibrinated blood. Fibrinolysis

If blood is kept constantly stirred or is whipped by means of a faggot of fine twigs or wires, the fibrin gradually collects upon their surfaces. In this way all the fibrinogen in a short time becomes converted into fibrin which is removed as it is formed. The fluid which remains is composed of serum together with the red and white cells and resembles blood in appearance and consistency, but of course is totally incapable of clotting. Defibrinated blood is frequently used in physiological experiments instead of employing an anticoagulant. Though defibrinated blood closely resembles ordinary blood it frequently contains toxic substances developed during the clotting process and is, therefore, unsuitable for transfusion into man. The formation of fibrin is actually hastened by the whipping action, since the greater surface to which the blood is exposed and the violent agitation to which the cellular elements are subjected cause greater destruction of platelets.

Blood which has been extravasated into the pleural, peritoneal or other body cavity is usually found to be fluid, if any considerable time has elapsed since the hemorrhage occurred. Such blood is found to be incoagulable, for it has already clotted and then undergone a natural process of defibrination, the clots having been dissolved as a result of the digestion of the fibrin (fibrinolysis) by proteolytic enzymes. The fluidity of menstrual blood is caused in the same way. A fibrinolysin, although not present in normal blood under ordinary circumstances, appears in plasma after shaking with chloroform.

A great deal of work has been carried out in recent

years on the fibrinolytic activity of the blood, and much has been learned of the mechanisms involved. Fibrinolysis occurs in a number of pathological conditions. It has been found, for example, that blood clots removed from a body immediately after sudden death undergo rapid dissolution due to the development of fibrinolytic activity. Also, blood taken from patients immediately after surgical operations, or from those who have suffered severe burns or hemorrhage, contains fibrinolysin, sometimes in high concentration. Through the agency of fibrinolysins the fibrinous exudate filling the alveoli of the consolidated lung in lobar pneumonia is dissolved preparatory to its absorption. The blood normally contains a proferment, which has been called *profibrinolysin* or *plasmin* (i.e., fibrinolysin in an inactive form), which under certain circumstances may be converted to the active enzyme by a second principle derived from the tissues. This is known as *fibrinokinase*. Activation can also be brought about artificially, as mentioned above, by chloroform and also by certain strains of streptococci, due to their possessing an activating enzyme, known as *streptokinase*. The latter substance together with *streptodornase* (*desoxyribonuclease*), an enzyme also produced by streptococci, and which hydrolyses desoxyribonucleoprotein of cell nuclei, has been employed in surgery to liquefy thick pus, and thus to render it more amenable to drainage (Ch 48).

Substances which hasten the clotting process (hemostatics)

Adrenaline when injected hastens the clotting process. Emotional excitement, muscular exercise and hemorrhage act similarly, probably through the liberation of the hormone from the adrenal medulla. Adrenaline has no effect upon the coagulation of blood after it has been shed, which suggests that it does not act directly, but through an intermediary, possibly another endocrine secretion.

The effects of thrombin and thromboplastin have been discussed. Tissue extracts especially those of lung and thymus which are rich in thromboplastin are powerfully coagulant, as are also the venoms of some species of snakes. The coagulant property of such venoms depends upon their containing a proteolytic enzyme which converts prothrombin to thrombin, the conversion occurs in the absence of ionized calcium.

Thrombin sprayed upon the bleeding surface in conjunction with fibrinogen to form a covering film or foam of fibrin, is an effective means of arresting bleeding, especially from numerous small vessels. Among other hemostatic agents used to hasten the clotting process are sodium alginate, derived from seaweed, and a gauze made of oxidized cellulose, which swells when soaked with blood. When sodium alginate comes into contact with the blood and serum of the wound, it is converted to calcium alginate, which "clots" to form a tenacious layer.

TESTS EMPLOYED IN THE INVESTIGATION OF DEFECTS IN THE CLOTTING MECHANISM

Coagulation time This is the time which the blood takes to clot after it has been shed. Obviously, any condition which decreases the coagulability of the blood lengthens the coagulation time and vice versa.

A number of methods have been devised to determine the precise moment when clotting occurs. The clotting time as determined by different methods varies considerably since the index or criterion of coagulation is not the same in all, and the conditions to which the blood is subjected are also different. On this account the values are not absolute and the results obtained by different methods cannot be compared strictly with one another.

A simple but rough method when a considerable quantity can be secured is to collect the blood into a small test tube and take as the coagulation time, the period elapsing from the moment the blood was shed to when it congeals, as indicated by tilting the tube from time to time. This procedure is employed in the Howell and the Lee-White methods. In the latter, 1 cc of blood is collected in a thoroughly cleaned Wasserman tube about 8 mm in diameter. The coagulation time of normal human blood as determined by this method is from 6 to 15 minutes.

A more accurate method and one which requires only a drop of blood is the following. The blood is drawn into a capillary glass tube about 4 or 5 inches long. A section of the tube is broken off from time to time. The time elapsing from the moment when the wound was made to that when fine threads of fibrin appear between the ends of the broken sections of the tube is taken as the coagulation time. The normal coagulation time by this method is about 5 minutes. There are many other methods for determining the coagulation time, e.g., Brodie's, Gibbs', Cannon and Mendenhall's, etc., but all have their defects and the foregoing is probably just as accurate as more elaborate procedures.

Prothrombin time, determination of the prothrombin concentration Quick's method is as follows. Nine volumes (about 2 cc) of blood obtained from a vein in a silicone-coated syringe are decalcified by the addition of 1 volume of 0.1 mol solution of sodium oxalate. After centrifuging, 0.1 cc of the oxalated plasma is mixed with 0.1 cc of thromboplastic material (prepared from fresh rabbit's brain) and recalcified with 0.1 cc of 0.025 mol solution of calcium chloride. The time from the addition of the calcium to when the first fine mesh of fibrin filaments appears is recorded by a stop watch. The test is carried out in a water bath at 37.5°C. The

prothrombin concentration in per cent of the normal is calculated as follows

$$\text{Prothrombin concentration} = \frac{K}{c t - a}$$

(per cent of normal)

$c t$ = clotting time, K , a constant = 303, a , a second constant = 8.7. Thus if the clotting time is 21 seconds the prothrombin concentration is $303/(21 - 8.7) = 25$ per cent of normal (approx.)

Clot retraction time This is measured by collecting a few cubic centimeters of blood into a test tube. After clotting has occurred the clot is separated from the walls of the tube by means of a fine wire and the tube placed in an incubator set at a temperature of 37°C . The clot of blood, normally, retracts to a firm mass, the serum separating out within a couple of hours. In certain types of purpura (thrombocytopenic) the clot remains bulky, soft and friable. In hemophilia, on the contrary, when clotting does occur, the clot usually retracts normally.

The bleeding time The determination of the bleeding time simply consists in pricking the skin and noting the time the drop of blood takes to clot sufficiently in order to close the puncture in the skin and stop the bleeding. The precise moment when bleeding ceases is determined by touching the blood from the tiny wound every few seconds with a piece of filter paper. The moment when the latter ceases to be stained is taken as the end point. The normal bleeding time is about $2\frac{1}{2}$ minutes.⁴ One might suppose that the bleeding time would be a better gauge of the body's ability to protect itself against hemorrhage than the coagulation time, but the fact that it is normal, as a rule, in hemophilia (see below) shows that this is not so. The ability to seal a wound in a minute vessel is dependent more upon vascular factors, contraction of the vessel wall and "glueing" together of endothelial surfaces, than upon the clotting of the blood. Actually, there is little relationship between the bleeding time and the coagulation time. The former is prolonged in conditions associated with low platelet counts. It is shortened by the local application or intravenous injection of platelet extracts, by the application of solutions with a low pH (around 5), or by the injection of posterior pituitary extracts.

⁴ Ivy and his associates recommend first raising the venous pressure in the arm by constricting it above the point where the skin puncture is to be made. A constricting pressure of about 40 mm Hg is applied. Tocantins has devised an instrument which makes an incision of approximately uniform length and depth.

HEMORRHAGIC DISEASES

It is well to emphasize here that a low serum calcium is never a cause of hemorrhagic disease. In hypoparathyroidism, in which the serum calcium may be depressed to less than half the normal value, the coagulation time is not lengthened.

Hypoprothrombinemia

Depression of the prothrombin level in the blood—*hypoprothrombinemia*—occurs under several conditions, such as vitamin K deficiency and severe liver damage. A bleeding tendency is not shown, however, according to Quick, until the prothrombin concentration falls to about 20 per cent of the normal. Other observers, however, place the critical level somewhat higher.

Hemorrhagic disease of the newborn has been shown quite definitely to be the result of a low prothrombin concentration. At birth and up to 6 hours thereafter, the prothrombin level is not far below normal, but apparently the baby comes into the world with a small reserve. This tends to become quickly exhausted, so that the prothrombin concentration may reach a low level at the end of 24 hours. This physiological hypoprothrombinemia is due, mainly, to lack of vitamin K. When the infant commences to take food, bacteria are introduced into the intestinal tract which, acting upon the contents of the intestine, synthesize the anti-hemorrhagic vitamin (ch. 55) and the hypoprothrombinemia is corrected automatically. Should the hemorrhagic state actually develop it is quickly arrested by the administration of vitamin K. Large doses may be given to the mother in the later months of pregnancy as a preventive. Hypoprothrombinemia and a hemorrhagic tendency also occur in certain diseases, e.g., acute yellow atrophy, associated with severe liver damage, in certain intestinal diseases, e.g., sprue and ulcerative colitis, and in the hemorrhagic disease of farm animals caused by eating spoiled sweet clover. In the animal disease the low prothrombin level is due to a toxic substance (*dicoumarol*, p. 118). The mode of action of this compound is not explained by any damaging effect which it might have upon the liver for in therapeutic doses it appears to have no such action though in large doses it may have. In the latter instance its anticoagulant action is then enhanced as a result of the interference with prothrombin manufacture by the liver. It appears to antagonize the action of vitamin K, being more

effective when the concentration of this vitamin is low. It may, therefore, be classed with the anti-vitamins.

The bleeding tendency which for years has been recognized to be a feature of obstructive jaundice, or when bile is lost to the exterior through a biliary fistula, has been shown to be the result of vitamin K lack. The deficiency is not due, however, to any dearth of the vitamin in the intestinal tract, but simply to the absence of bile salts without which the vitamin cannot be absorbed into the blood stream (p 541).

Hypoprothrombinemia may be caused by bacteriostatic agents, e.g., sulfonamides active against the intestinal flora responsible for the synthesis of vitamin K.

Hemophilia

The coagulation time but not the bleeding time as defined above is greatly prolonged in this condition. In some cases the blood after removal from the body may show no signs of clotting after an hour or so. On this account a fatal hemorrhage may follow a wound that would be trivial in a normal person. The extraction of a tooth, a slight accident or a minor operation have resulted in death on many occasions. The subjects are known popularly as "bleeders." Hemophilia is transmitted as a sex-linked recessive character. The males are affected but do not transmit the disease, the females, though they transmit the disease do not suffer themselves (law of Nasse (1820)). A father, then, who is a "bleeder" does not transmit the disease to his children in a manifest form. His sons are entirely free from any taint but his daughters inherit the disease in a masked or latent form and transmit it to their offspring. Of these the males show the disease, and the daughters again are not "bleeders" but are "carriers" for the next generation. In other words the manifest disease skips a generation. Some of these bleeder families have been traced back for a hundred years or so and cover in some cases 5 generations and hundreds of individuals. The incidence of the disease in these instances followed in general the plan outlined.

Not every person with a tendency to bleed can be classed as a hemophiliac. True hemophilia answers to the following criteria:

- (a) Males only
- (b) History showing characteristic type of inheritance
- (c) Coagulation time of drawn blood, but not the bleeding time, is greatly delayed
- (d) No reduction in platelets, but these show lessened fragility

above. It is theoretically possible, should a "bleeder" and a "transmitter" marry, for the female as well as the male children to be "bleeders." Such a coincidence is so very rare, and the cases of true hemophilia in the female, therefore, so few, that it is usually stated that hemophilia is confined to the male sex.

The defect in the clotting mechanism of hemophiliacs is not definitely known, but a factor seems to be a lack of a sufficient amount of active plasma thromboplastin. The plasma thromboplastin may be actually deficient in amount, fail to undergo conversion from an inactive to an active (Quick) form, or is inactivated or antagonized in some way by inhibitory substances (Tocantins). Tissue thromboplastin or normal platelet-free plasma added to hemophiliac blood restores the normal clotting time. Thus, there is no lack of prothrombin, nor is heparin responsible for the prolongation of the clotting time. The platelets in hemophilia are not reduced in number, but they are less fragile than normal. Yet, platelet stability is not the cause of the bleeding in hemophilia. The platelets, contrary to a previous belief, do not contain significant amounts of thromboplastin so that they do not, through their failure to disintegrate readily, withhold this principle from the plasma. Moreover, complete disintegration of the platelets in hemophiliac blood by saponin (Pickering) does not correct the defect in the clotting mechanism, and Quick found that platelets from the blood of a hemophiliac restored to normal the clotting time of blood which, due to a deficiency of platelets, clotted slowly. Quick draws the conclusion from his experiments that the fundamental difference of hemophiliac from normal blood is a congenital deficiency of plasma thromboplastinogen (p 114).

It may seem surprising that the bleeding time is not prolonged in hemophilia, but it must be remembered that the wound which is made in determining the bleeding time is very small, and even though the clotting mechanism itself is defective, is capable of being sealed by the contraction of the walls of the minute vessels, and the sticking together of the apposed endothelial surfaces. It is when larger vessels are opened that persistent hemorrhage occurs in hemophilia.

Logically, the measure to be used to arrest bleeding in hemophilia, and the one that has proved most effective, is to furnish the missing thromboplastic material either by blood transfusion or by the application of tissue extract to the bleeding point.

Barnet and Macfarlane have reported a series of cases in which highly successful results were obtained from the application of snake venom, which is rich in thromboplastin, to the wound A 1 in 10,000 solution of venom of Russell's viper was used

Parahemophilia This is the name given by Owren to a very rare hemorrhagic state caused by the lack of his factor V (labile factor or Ac-globulin, p 114)

Two other rare hemorrhagic conditions are *heparinemia* and *afibrinogenemia* Heparinemia, as its name indicates, is the presence of heparin in the blood This has been encountered after extensive X-radiation of either animals or man or following the administration of nitrogen mustards for leukemia The blood may be rendered quite incoagulable. The bleeding is unaffected by vitamin K administration, but may be checked by protamine or toluidine blue given intravenously As already mentioned (p 118), the experimental production of peptone or anaphylactic shock is associated with the discharge of heparin into the circulation Oddly enough, tumours of mast cells which contain relatively enormous quantities of heparin do not affect the coagulability of the blood A congenital absence or very low concentration of fibrinogen—*afibrinogenemia*—is a very rare anomaly of the blood-clotting mechanism

Purpura This term is applied to a variety of hemorrhagic states in which spontaneous bleeding occurs beneath the skin, from the mucous membranes or into joints The subcutaneous hemorrhages appear as small or large purplish spots (*petechiae* and *ecchymoses* respectively) which gradually pass through the color changes characteristic of a bruise Since purpura is so varied in its characters and occurs in diseased states that are so widely different, it should be considered as a symptom of a disorder of the blood-vascular system rather than as a disease in itself It occurs in the malignant forms of many acute diseases—smallpox, scarlet fever, diphtheria, streptococcal infections, etc Subcutaneous and submucous hemorrhages are also features of scurvy, leukemia, certain anemias and of the action of various toxic agents, e g, snake venoms, drugs and chemicals In some types of purpura (e g, infectious and toxic forms) *deterioration of the capillary wall* probably plays the chief rôle—the red cells escaping into the subcutaneous tissues or from the mucous membranes through capillary defects On the other hand a

great *reduction in platelets* (thrombocytopenia) appears to be an important if not the essential defect in other types and a condition resembling purpura hemorrhagica (see below) has been induced in animals by the injection of an antibody developed for the destruction of platelets Nevertheless the experiments of Bedson suggest that reduction of platelets or even their entire absence is not capable alone of inducing purpura Some injury to the capillary wall must exist as well This observer also found that splenectomized guinea pigs were highly resistant to the action of the anti-platelet serum A theory has been advanced which causally relates the capillary defects to the thrombocytopenia It is probable that the platelets serve, normally, to protect the capillary wall and, by their deposition upon the endothelium, act as a seal against the escape of red cells through weakened points which are constantly occurring from general wear and tear The reduction in platelets may result from several causes There may be increased destruction of these elements or the normal mechanism of their production may be interfered with, as in aplastic anemia and leukemia

The *coagulation time* in purpura is usually within normal limits but the *bleeding time* is as a rule prolonged The clot which forms in the blood after it has been shed is said to be softer than the normal and does not contract and express the serum in the usual way

THE CAPILLARY RESISTANCE TEST When cutaneous purpuric spots do not occur spontaneously they may be induced in susceptible persons (latent purpura) by means of a tourniquet or blood pressure cuff applied to the upper arm so as to obstruct the venous return but not the blood flow in the artery The obstruction is maintained for 5 minutes In scurvy and other conditions associated with weakness of the capillary membrane the increased intracapillary pressure so induced results in the formation of small hemorrhagic points (*petechiae*) beneath the skin of the forearm A more precise method consists of the application to the skin (usually of the forearm just below the antecubital fossa) of a small suction cup and determining the minimum negative pressure which, when applied for one minute, is required to produce petechiae In health this lies between -200 and -300 mm Hg A minimum effective pressure less than -200 mm Hg is abnormal

Purpura hemorrhagica or Werlhof's disease (thrombocytopenic purpura) is associated with a great reduction

in the platelet count. They may be almost absent from the circulation. Splenectomy is followed by an increase in platelets and frequently effects a cure. Recently snake venom has been successfully employed in the treatment of this disease. The observations of Troland and Lee point to the spleen in this disease as the source of a toxic material having a destructive action upon platelets. An acetone extract of the spleens of patients suffering from thrombocytopenic purpura when injected into rabbits caused a progressive fall (from 640,000 to 20,000 in 24 hours) in the platelet count. They have given the name *thrombocytopen* to the unidentified active principle in the extracts. Extracts prepared in the same way from normal spleens were without effect upon the platelet count. An abnormal number of megakaryocytes in the bone marrow especially of younger forms has been reported in purpura hemorrhagica and it is suggested that the failure of these to produce platelets is the fundamental fault.

In one form of purpura—*purpura fulminans*—the hemorrhages are exceptionally severe and the subcutaneous extravasations often extensive. Death may occur within a few days from loss of blood. In this state there is undoubtedly an affection of the capillary endothelium as well as reduction in the number of thrombocytes. Purpuric manifestations may occur with more or less severe symptoms of gastro-intestinal irritation—*Henoch's purpura*—or in others bleeding into the joints (*Schonlein's disease*) is a prominent feature. In either of these forms the purpuric spots may occur in association with urticarial wheals or more generalized edematous swellings of the subcutaneous tissues. Capillary damage is probably either the essential or a contributory factor in these states.

Calcium is commonly employed in the treatment of the various types of purpura and appears to reduce the hemorrhagic tendency but the manner in which the effect is brought about is obscure, it is not through increasing the coagulability of the blood.

INTRAVASCULAR CLOTTING—THROMBOSIS

Coagulation within the vessels may be brought about experimentally (a) By the injection of thrombin into the blood stream (see p. 116) Repeated injections of small amounts of thrombin cause the blood, after the first immediate increase of coagulability, to become for a few hours less coagulable than normal (negative phase). This is due apparently to the compensatory production of an antithrombin. No permanent reduction in coagulability can however be produced by repeated injections of thrombin over a period of several weeks. (b) By the rapid injection of a tissue extract (thromboplastin) particularly of the lung, thymus or lymph glands. The repeated injection of small amounts of tissue extract may have the reverse

effect—decreased coagulability. This was shown by Mellanby and later by Mills and his associates to be the result of the gradual deposition of fibrin upon the vascular walls. The blood failed to clot simply because it had been depleted of fibrinogen. (c) By injury to the vessel wall either by chemical, mechanical or infective agencies, a roughened surface being thereby exposed to the blood stream. Thromboplastin is also liberated from the injured vascular wall.

The formation of a compact and solid mass of blood in a vessel (vein or artery) which partially or completely occludes its lumen is spoken of as *thrombosis*. The plug itself is termed a *thrombus*. A thrombus may be formed simply by the clotting of the blood in the usual way, as occurs in a grossly damaged vessel. On the other hand the cellular elements of the blood may form a solid mass within a vessel and block its lumen independently of the clotting process. This occurs sometimes in the smaller vessels when the red cells become agglutinated into large clumps (agglutinative thrombus) as after the transfusion of incompatible blood, or when the blood forms a sludge (p. 10). Once the vessel has become plugged in this way a true clot forms eventually in the stagnant column of blood behind the block. Thrombosis usually occurs in a vein, rarely in an artery.

When a thrombus, or a portion of it, becomes detached and carried away in the blood stream to become impacted in a vessel at some remote point in the circulation, it is spoken of as an *embolus*. When arising from a venous thrombus the embolus is likely to lodge in a vessel of the lung. If formed in the left heart it may plug a cerebral artery. The main vessel leading to a circumscribed area may become obstructed and the anastomosing channels be insufficient to maintain the nourishment of the tissue. The isolation of an area in this way by the obliteration of its artery is spoken of as *infarction*, and the necrotic area as an *infarct*.

Causes of thrombosis in man

The causes of thrombosis in the human subject may be grouped under the following heads

(1) **INJURY TO THE VESSEL** Complete mechanical obstruction of the circulation as by a ligature or pressure upon a vessel from without does not alone cause thrombosis. The latter is usually associated with some injury to the inner coat of the vessel, unless this occurs the blood may remain fluid for a considerable time. A

thrombus, often very small in size, may form later merely as the result of the slowing of the stream, the formation of eddies in the blind pocket and the deposition and lysis of platelets. Septic infection is one of the commonest causes of endothelial damage leading to thrombosis. Some degree of thrombosis is always associated with acute inflammatory processes of septic origin. The thrombosis may be strictly localized to the smaller vessels at the site of the infection, or may extend into larger venous trunks, and is then due to the spread of the infection along the vein walls (phlebitis). The thrombosis of the femoral vein during the puerperium or of a cerebral sinus following mastoid disease are typical examples of infective thrombosis.

(2) **AGGLUTINATION OF CORPUSCLES** (agglutinative thrombus, p. 44)

(3) **TOXIC THROMBOSIS** Certain chemical poisons, e.g., arsenical compounds, mercury, potassium chlorate, etc., may cause intravascular clotting. Poisonous mushrooms, certain snake venoms, as well as toxins formed within the body, as in eclampsia, or extensive burns, may induce thrombosis. The manner in which these various agencies act, whether by injury to the vessel wall, disintegration of blood elements, or through their effects upon some phase of the clotting process is obscure.

(4) **SPONTANEOUS OR POST OPERATIVE THROMBOSIS**
PULMONARY EMBOLISM. Spontaneous thrombosis may occur under a variety of conditions, and in practically any vein in the body. Thrombosis, particularly of the veins of the lower limbs, not infrequently follows operations upon the abdominal or pelvic organs. When it occurs under these circumstances it is always a cause of anxiety to the surgeon because of the danger of the clot becoming detached and carried to the lung where it may block a branch of the pulmonary artery (pulmonary embolus). Infection is frequently blamed for the thrombosis, but in most cases no evidence for this can be found. Also, the point where the thrombus forms—usually in the deep veins of the lower limb, such as the plantar or calf veins, or the femoral vein—lies at a distance from the field of operation and is separated from it by a considerable extent of healthy tissue. This precludes direct spread either of infection or of thrombus formation from the site of the wounded tissues.

The principal factors concerned in spontaneous thrombosis, especially following operation, will now be discussed.

(a) *Endothelial damage* Some injury or a slight alteration in the vascular lining which makes it "wettable" probably initiates the process.

(b) *Slowing of the blood stream* as a result of enfeebled heart action, prolonged confinement to bed, debilitating diseases associated with a low metabolic rate and hypotension, or immobility of the limbs, seems to be an important factor in causation. It encourages the deposition of platelets on the vascular wall, and may be accompanied by "sludging" of the blood (p. 10).

(c) *Deposition of platelets* Aschoff, some years ago, made an experimental study of postoperative thrombosis and gave an interesting and logical explanation of the processes involved. He states that when the vein is completely obstructed, as is usual, the thrombus has a white portion or head which is directed proximally (toward the heart) and a dark red portion or tail. If the flow of blood is not completely blocked the thrombus consists of the white portion alone. A minute examination reveals a framework of ribs or beams extending from the wall of the vein and traversing the entire substance of the white thrombus. The ribs are made up chiefly of massed platelets covered with a layer of leukocytes. *There is little or no sign of fibrin* or of red cells, this white plug is not a clot in the ordinary sense. The longer red portion or tail of the thrombus extends distally for a variable distance along the vessel and is made up of all the elements of the blood in their normal proportions. Fibrin threads are plentiful. This portion is evidently formed by the coagulation of the blood en masse. The mode of deposition of the platelets is compared to the manner in which sand, though kept in suspension in a rapid stream, is deposited in a ribbed pattern upon the sea shore or at a river's mouth where the current is slowed. Though the flow of blood in the veins is continuous it is not absolutely even, eddies occur in the venous current and even slight muscular movement causes a certain irregular wave like motion of the blood column which accounts for the deposition of the platelets, not in a continuous even layer but in ripples. Also, owing to their lower specific gravity the platelets leave the axis or core of the stream and separating from the other blood elements come to occupy the more slowly moving zone next the vessel wall. Finally, it is the reduction of the velocity in this outer zone of the current to a certain critical level that causes the platelets to settle upon the vascular walls. The ribs or ridges increase in height by the aggregation of fresh platelet masses, and secondary ridges are later formed upon the primary ones until at last the fabric extends like a coral growth into the axis of the stream.

In the view of later investigators, some change in the endothelial lining must occur to render it "wettable", (a) above, before platelets will adhere and undergo lysis.

In the auricles during auricular fibrillation, in the slowing of the stream caused by a local dilatation of the blood channels, e.g., in an aneurysm or varicose vein, and wherever eddies are produced, a white thrombus either with or without a covering of clotted blood, is likely to form as a result of platelet deposition.

The observations of Sandison are pertinent to the question of thrombus formation. Examining the blood flow through the capillaries of the rabbit's ear by the transparent chamber method, he observed thrombi in the process of formation. When the blood flow was retarded platelets were seen to cling to the wall of the capillary or venule. By the deposition of successive

layers, a white thrombus was formed which as it grew extended a considerable way along the blood channels. A small proportion of leukocytes were included in the mass, but no erythrocytes.

(d) *Blood changes* The platelets are increased after operations, and these elements show a greater tendency than is normal to clump together. The greater stickiness is due, according to Wright, to newly formed cells discharged from the bone marrow. The fibrinogen also is increased and, as a result of this, the rate of sedimentation of the corpuscles is hastened (p. 66), clot retraction is usually more pronounced than normally.

The *coagulation time* is shortened for the first few days after operation but it is doubtful whether an altered relationship between the different factors concerned in the clotting mechanism itself is of prime importance in the production of post-operative thrombosis. At first sight the liberation of thromboplastin from the damaged tissues would seem to be a probable cause of the thrombosis. Yet if this were responsible one would expect that thromboses would occur immediately following operation and not as is actually the case a week or ten days later. Moreover, thrombi composed of masses of platelets have been induced in animals whose blood had been rendered incoagulable, and post-operative thrombosis may occur in subjects whose blood shows a clotting time within the normal range.

Though some alteration in the vascular endothelium and slowing of the venous blood flow are probably the most important factors in the production of post-operative thrombosis, alterations in the blood itself which follow tissue injury, namely, the increase in the number of platelets and the rise in fibrinogen concentration, no doubt encourage the formation of the white thrombus. Anhydremia by increasing the viscosity of the blood may also favor its occurrence.

Resume

The most probable sequence of events in the formation of a thrombus postoperatively is as follows. First, some abnormality occurs in the endothelial lining of a vessel and its unwettable property is abolished. Platelets adhere and agglutinate at this point and a white thrombus forms. Platelet accumulation and the growth of such a white thrombus are favored by a sluggish blood flow. As the platelets disintegrate, the series of reactions leading to thrombin production and the formation of a fibrin net (with or without entrapped red cells) is initiated. The platelet thrombus with the clot subsequently formed may adhere

firmly to the vascular lining, in which event the thrombus remains fixed and, though it may occlude the vessel, there will be relatively little danger of its being swept away in the blood stream to block a distant vessel. Quick attaches much importance to clot retraction as a factor to the further growth of the clot and the production of an embolus. If clot retraction is pronounced, as in anemia or when the platelet count is high, thrombin is expressed from the interstices of the fibrin net and causes extension of the clot. The latter, attached by its base to the endothelium, but with its other end floating in the circulation and pointing in the direction of the blood flow, is continually adding to its length by fresh accretions of clotted blood. This elongated mass has a rather precarious hold and after a time, loosened by the force of the blood stream or perhaps by some sudden movement of the patient, is detached and carried away. According to this conception, clot retraction which draws the thrombus from the vascular wall (except at its base to which for a time it is moored) favors its enlargement in the blood stream, and is an essential factor in its ultimate detachment as an embolus. The abnormality in the vascular lining which starts the train of events leading to thrombosis can not always, nor perhaps often, be demonstrated, even histologically, but from indirect evidence some such alteration is believed to be present almost invariably.

The following are the chief measures which have been employed in attempts to prevent post-operative thrombosis.

(1) Early movement of the limbs, to favor the venous flow.

(2) Avoidance of any restriction to respiration—which is an important factor in aiding the venous return from the limbs.

(3) Thyroid extract administration, to raise the metabolism and increase the circulation rate.

(4) Plenty of fluids, to prevent dehydration, and a diet composed largely of carbohydrate since the platelet count and the fibrinogen are raised by a high protein diet.

(5) The treatment of any anemia which may exist.

(6) The use of anticoagulants such as heparin by continuous intravenous administration (p. 116) or of dicoumarol by mouth, which tend to prevent platelet agglutination.

CHAPTER 13

THE MECHANISMS REGULATING THE REACTION OF THE BODY FLUIDS

PHYSICO-CHEMICAL PRINCIPLES

THE ELECTROLYTIC DISSOCIATION THEORY

When certain chemical substances, such as acetic acid, hydrochloric acid and many others, are dissolved in water a proportion of their molecules—the proportion varies with the particular substance—undergo dissociation into their constituent ions. The latter move through the solution in all directions, some collide with one another and recombine, while a corresponding number are produced by the dissociation of other molecules. By the junction of some ions and the separation of others, the balance between undissociated (unionized) and dissociated (ionized) molecules is kept constant. That is, equilibrium is established. Substances whose molecules are dissociated in this way are known as *electrolytes*. An ion is an electrically charged atom or group of atoms and is referred to as a *cation* or an *anion* respectively according to the nature of the charge—positive or negative—which it holds. Cations are denoted by a plus sign or simply by a dot placed above the atomic symbol. The negative ion (anion) is indicated by a minus sign or an oblique dash. Thus HCl dissociates into the ions H^+ and Cl^- , H_2O itself into H^+ and OH^- and HCO_3 into H^+ and HCO_3^- . When an electric current is passed through the solution each ion carries a charge of electricity to one or other electrode, the anions (negative ions) migrating to the positive electrode or anode and the cations (positive ions) to the cathode. The conduction of an electric current through the solution will therefore be influenced by the concentrations of ions in the solution (i.e., upon the degree of dissociation of the particular electrolyte). Solutions of such substances as sodium chloride are good conductors, while solutions of others like cane sugar which undergo little or no dissociation are not much better conductors than pure water.

When (1) the degree of dissociation of an acid or other electrolyte and (2) the total concentration of the acid in the solution are known, then the concentrations of dissociated and undissociated molecules can be calculated. It is found that after equilibrium has been established the product of the concentrations of ions divided by the concentration of the undissociated molecules gives at a given temperature and for a given electrolyte, a constant value which is known as the *ionization* or *dissociation constant*. This represents the operation of the mass law¹ as applied to the dissociation of

electrolytes. So if HA represents an acid which is dissociated into cations H^+ and anions A^- then K_a represents the dissociation constant of the acid. The brackets enclosing the letters represent molar concentrations. Thus

$$\frac{\text{Dissociated}}{\text{Undissociated}} = \frac{[H^+] \times [A^-]}{[HA]} = K_a \text{ (Ionization constant)}$$

After equilibrium has been reached the velocity at which the molecules HA dissociate into H^+ and A^- is equal to that at which the latter combine to form HA. In this way an equilibrium between dissociated and undissociated molecules is maintained, thus, $HA \rightleftharpoons H^+ + A^-$. For weak electrolytes the degree of dissociation increases with dilution, so that at infinite dilution the dissociation is complete. The greater the degree of dissociation the higher will be the numerator of the equation, and the higher consequently will be the value of the dissociation constant.

HYDROGEN ION AND HYDROXYL ION CONCENTRATIONS

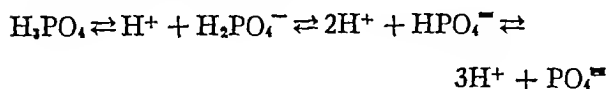
The ionization of acids

If a measured amount of a normal solution of acetic acid be titrated against a normal solution of sodium hydroxide a definite amount of the base will be required for neutralization. If the same quantity of a normal solution of HCl be taken and titrated as before, it will be found that precisely the same amount of base will be required for complete neutralization. From this it might be assumed that the two acids are equally strong. But it is known that at similar concentrations the acid properties of hydrochloric are incomparably greater than those of acetic acid. The former inverts cane sugar more powerfully, it tastes more acid, it has a greater destructive effect upon animal tissues, and it will displace acetic acid from its salts.

Hydrochloric acid is dissociated into hydrogen ions and chlorine ions. To the H^+ -ion it owes its acid properties, it is the hydrogen ion concentration—reacting substances. This law while it holds for dilute solutions of such weak electrolytes as are present in blood is inaccurate in the case of strong electrolytes in concentrated solutions.

¹ The mass law states that the velocity of a chemical reaction is proportional to the concentrations of the

tion and not the total number of hydrogen atoms in its molecule that is responsible for the acid characters of any acid. The greater the degree of dissociation of the acid, the greater is the number of H-ions which it will yield in solution, i.e., the greater will be the H-ion concentration, and the greater consequently will be its acid nature. Thus HCl though it contains only one atom of hydrogen in its molecule undergoes almost complete dissociation and is in consequence a much stronger acid than carbonic (H_2CO_3), whose molecule contains two hydrogen atoms but dissociates to a very small extent into H^+ and HCO_3^- , carbonic acid dissociates in two stages. Phosphoric is another weak acid which contains three atoms of hydrogen. It dissociates, however, to a greater extent than carbonic acid and in three stages



The greater the degree of dissociation of an acid the higher will be the value of its dissociation constant. The latter therefore is a true measure of "acid properties." The dissociation constant of a very weak acid such as carbonic is only 0.0000003 (first stage), the first dissociation constant of phosphoric acid is 0.011.

The ionization of bases

The OH ion on the other hand gives to bases in aqueous solution their characteristic properties, and when a base and acid neutralize one another it is the union of the hydrogen ion of the acid with the hydroxyl ion of the base with the formation of a molecule of water that brings about neutralizations, thus



The ionization of a base may therefore be represented by an equation analogous to that given for an acid. Thus

$$\frac{\text{Dissociated}}{[\text{B}^+] \times [\text{OH}^-]} = K_b$$

Undissociated
[BOH]

If the hydrogen ions are in excess the solution is acid, if the hydroxyl ions are in greater concentration the solution is alkaline.

The ionization of water—The reaction of a solution expressed in terms of hydrogen ion concentration

The dissociation of water may be represented by the equation

$$\frac{[\text{H}^+] \times [\text{OH}^-]}{[\text{H}_2\text{O}]} = K_w$$

Pure water has a definite though very slight conductivity value. The concentration of the molecules that are dissociated into H^+ and OH^- as compared with those undissociated is almost infinitesimal, the concentration of the H^+ and OH^- is so small that their presence produces no measurable decrease in the concentration of the total water molecules. The equation then may be written simply—

$$[\text{H}^+] \times [\text{OH}^-] = K_w$$

K_w , the dissociation constant for water therefore represents the product of these ionic concentrations and amounts to no more than 0.00000000000001.

Expressed more briefly, though perhaps more cryptically, the value is 1×10^{-14} . The symbol -14 to the right of the figure 10 is termed the negative exponent or index, and means that in order to express the value in the form of a decimal fraction, as in the preceding paragraph, the figure 1 must be placed 14 places to the right of the decimal point. Expressed as a vulgar fraction 10^{-14} would be $\frac{1}{100000000000000}$.

Other examples of this system of notation are

$$10^{-1} \text{ means } 0.1 \text{ or } \frac{1}{10}$$

$$10^{-2} \text{ means } 0.01 \text{ or } \frac{1}{100}$$

$$10^{-3} \text{ means } 0.001 \text{ or } \frac{1}{1000} \text{ and so on}$$

We already know that the H^+ and OH^- must be present in equal concentration and that the dissociation constant 1×10^{-14} is the product of these concentrations, i.e., the square of either. Therefore the square root of 1×10^{-14} i.e. 1×10^{-7} indicates the concentration either of the H^+ or of the OH^- .

The actual weight of ionized water in 1000 cc (1 liter) of water is therefore

$$1 \times 10^{-7} \text{ grams H ions (atomic weight of hydrogen = 1)}$$

$$17 \times 10^{-7} \text{ grams OH ions (atomic weight of oxygen = 16)}$$

$$\text{Total } 18 \times 10^{-7} \text{ grams ionized H}_2\text{O}$$

A solution containing 1 gram of hydrogen ion per liter is known as one having a normal concentration of hydrogen ions. Seventeen grams of OH ions in the same quantity of fluid is known as a normal solution of hydroxyl ions. Water is therefore 10^{-7} normal in both H^+ and OH^- ions. Since both are in equal numbers per unit of fluid, water is neutral in reaction.

If hydrochloric acid be added to water it is dissociated to the extent of over 90 per cent into H^+ and Cl^- . The concentration of H^+ in the water will therefore be greatly increased by the addition of acid. But it has already been seen that the product of the concentration of H^+ and OH^- in water is constant (K_w). This statement applies not only to pure water but to all aqueous solutions. *This is a fundamental fact* and upon it the determinations of the hydrogen ion concentration and the notations used for its expression are based. When therefore the H^+ are increased the OH^- must undergo a reciprocal reduction. If, for instance, the H^+ concentration is increased above its value in pure water, that is from 1×10^{-7} to say 1×10^{-6} (0.000001 N to 0.0000001 N), then the concentration of OH^- must undergo a reduction to 1×10^{-8} in order that the product of these two shall remain the same ($1 \times 10^{-6} \times 1 \times 10^{-8} = 1 \times 10^{-14}$). On the other hand if a base be added to water and the OH^- concentration increased from its value in pure water, i.e., from 1×10^{-7} to, let us say, 1×10^{-6} , the H^+ concentration must undergo a corresponding diminution from 1×10^{-7} to 1×10^{-8} (0.0000001 N to 0.00000001 N). Therefore all that is required in order to indicate the concentrations of both H^+ and OH^- ions is an expression which denotes the concentration of either one. The concentration of the H^+ has therefore been chosen to express the reaction of a solution. The term "hydrogen ion concentration" is abbreviated to the symbol cH. If the hydrogen ion concentration (cH) is greater than that of pure water the reaction is acid, if less than this value, the reaction is alkaline, and if precisely the same, the reaction is neutral. Thus

If cH equals that of pure water i.e. 1×10^{-7} the reaction is neutral.

If cH exceeds that of pure water e.g. from 1×10^{-7} to 1×10^{-6} the reaction is acid.

If cH is less than that of pure water e.g. from 1×10^{-7} to 1×10^{-8} the reaction is alkaline.

The values represent the H^+ concentrations in terms of a normal solution (i.e., one containing 1 gram of ionized hydrogen per liter). No matter

how concentrated in a solution the H^+ may be, some OH^- are always present, and vice versa, with maximum alkalinity there are always present a small number of H^+ .

The symbol pH

The value of the hydrogen ion concentration as indicated above was found inconvenient. When the cH was 1×10^{-6} , 1×10^{-8} etc., the first factor could, for the sake of simplicity, be omitted. The expression then became cH 10^{-6} , etc. But when the first factor was other than unity, e.g., as in 5×10^{-6} or 1.3×10^{-7} or 4×10^{-8} , this obviously could not be done. The writing of such expressions became cumbersome and the hydrogen ion concentrations which they represented were difficult to keep in mind.

Sorensen in 1909 introduced a system of notation by which the negative exponent of the common logarithm (i.e., to the base 10) of the decimal fraction expressing the H-ion concentration is employed as a positive number. For example the H-ion concentration of serum is 0.00000004 normal. The cH might be expressed either as 4×10^{-8} N, or as $1 \times 10^{-7.40}$ N. The latter expression is simplified by omitting the 10 and replacing the minus sign of the negative exponent by the symbol pH (hydrogen ion exponent). The value 7.40 is derived as follows. The log of the figure in the cH expression above, i.e., 4, is 0.6021 (cf log tables). This log subtracted from the negative exponent, i.e., 8, gives the required figure 7.40. Similarly a solution 0.0000005 normal in H-ions may be expressed either as cH 5×10^{-6} or as cH $1 \times 10^{-5.3}$. The log of 5 is 0.6990, the pH is, therefore, $6 - 0.6990 = 5.3$.

It must be remembered that a rise or fall in the pH indicates a change in hydrogen ion concentration (cH or $[H^+]$) in the opposite direction. Thus a pH of 5.00 indicates a higher acidity than a pH of 6.00, and a pH of 9.00, a lower hydrogen ion concentration than a pH of 8. pH 7.00 of course indicates neutrality. Attention should be drawn to another point that is not quite so obvious, namely that a solution of a pH 5.00 is more acid not simply by a sixth than a solution of pH 6.00, but is in fact ten times more acid. A pH of 3.00 represents an acidity 1000 times greater than a pH of 6.00. Similarly a solution of pH 11.00 is 1000 times more alkaline than one of pH 8.00. Recalling the decimal fraction which these figures represent will make this clear. Thus pH 5.00 = 0.00001 normal and pH 6.00 = 0.000001 normal.

TABLE 15

HCl ADDED	$\frac{\text{H}_2\text{CO}_3}{\text{NaHCO}_3}$	$[\text{H}^+]$	$[\text{OH}^-]$	RELATIVE ACIDITY	RELATIVE ALKALINITY
grams					
0	2 27 11 9	0 000000057 N	0 000000176 N	0 57	1 76
10	2 27 11 5	0 000000059	0 000000170	0 59	1 70
50	2 27 10 0	0 000000068	0 000000142	0 68	1 47
100	2 27 8 2	0 000000083	0 000000120	0 83	1 20
150	2 27 6 3	0 000000108	0 000000093	1 08	0 93
200	2 27 4 4	0 000000154	0 000000065	1 54	0 65
250	2 27 2 6	0 00000026	0 000000039	2 6	0 39
300	2 27 0 68	0 0000010	0 000000010	10	0 10
310	2 27 0 31	0 0000022	0 000000045	22	0 045
318	∞	0 00026	0 0000000039	260	0 0039
320	—	0 00045	0 0000000022	450	0 0022
330	—	0 0027	0 00000000037	2700	0 00037

BUFFERS

These are substances which when present in a solution maintain the latter at a relatively constant pH when an acid or alkali is added to it. That is, a buffer has the power to "soak up" or "tampon" the acid or base, it takes up the shock so to speak of the strong acid or base, hence the term buffer. A buffer or buffer system consists of two parts—a weak acid and one of the salts of that acid. Acetic acid and sodium acetate, carbonic acid and sodium bicarbonate, phosphoric acid and sodium phosphate, are a few examples of such buffer pairs. Solutions of buffers are used in physiological experiments when for any reason it is desired to maintain the fluid medium at a constant hydrogen ion concentration, as for example in the study of the action of ferments.

The hydrogen ion concentration of a buffer solution is proportional to the ratio of the concentration of the free acid to the concentration of acid bound to base. The ratio of the buffer pair may be expressed as follows:

Concentration of free acid

$$\frac{[\text{HA}]}{[\text{BA}]}$$

Concentration of bound acid

The hydrogen ion concentration of the solution is equal to the product of the dissociation constant of the acid (K) and the ratio of the buffer pair. Thus

$$\text{Hydrogen ion concentration } [\text{H}^+] = K \frac{[\text{HA}]}{[\text{BA}]}$$

In order to express the H^+ ion concentration in terms of pH, the value for K is converted by means of logarithms to the corresponding pK figure. The equation then becomes

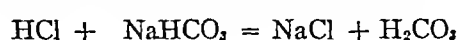
$$\text{pH} = \text{pK}_1 + \log \frac{[\text{BA}]}{[\text{HA}]}$$

This is known as the Henderson-Hasselbalch equation (see also ch. 32).

The effect of a buffer in preventing changes in reaction of a solution when acid is added may be illustrated by an example given by L. J. Henderson (table 15).

One hundred liters of a 1 per cent solution of sodium bicarbonate is made up and kept in contact with an unlimited atmosphere containing 0.1 per cent of carbon dioxide. Time is allowed for the solution to absorb CO_2 and for equilibrium to be established between the pressure of gas in the atmosphere and that in the fluid. The temperature is kept constant at 17°C . The solution to start with is slightly alkaline. Hydrochloric acid is added in successive amounts up to 330 grams. It may be seen from the accompanying table that the reaction of the solution is little altered after as much as 150 grams of HCl have been added. Even after the addition of 250 grams the H^+ -ion concentration is no more than 2.6 times that of a neutral solution. It is not until the bicarbonate has been completely used up that the acid exerts any great effect. The ratio $\frac{\text{H}_2\text{CO}_3}{\text{NaHCO}_3}$ remains almost unchanged until some 50 grams of HCl have been added.

The reaction may be expressed as follows:



In this reaction three things have occurred.

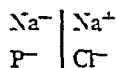
(1) The very strong acid, hydrochloric, has disappeared, the very weak acid, carbonic, having taken its place.

(2) The weak carbonic acid is volatile and is quickly got rid of by diffusion into the atmosphere. This continues until equilibrium between the pressure of gas in the atmosphere and that in the solution has been re-established.

(3) The bicarbonate has served as a reserve of base which has soaked up or buffered the added acid. The bicarbonate has of course been reduced and finally completely used up in the process.

THE DONNAN THEORY OF MEMBRANE EQUILIBRIUM

The Donnan effect is a state of ionic equilibrium set up between two sides of a membrane by the presence in a solution of electrolytes or an ion (such as protein) to which the membrane is impermeable. It is explained best by an illustrative example. Let two electrolytes, Na^+Cl^- and Na^+P^- , be separated by a membrane which is permeable to the ions Na^+ and Cl^- but impermeable to P^- . Thus,



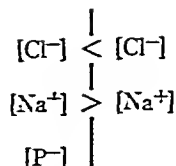
When equilibrium has become established the product of the concentrations of the *diffusible* ions (Na^+ and Cl^-) on one side of the membrane equals the product of the concentration of these ions on the other side.

$$\frac{[\text{Na}^+]}{[\text{P}^-]} = \frac{[\text{Na}^+]}{[\text{Cl}^-]}$$

Also on either side of the membrane the total concentration of anions (whether diffusible or non-diffusible) is equal to the concentration of cations, each solution being electrically neutral.

$$[\text{Na}^+] = [\text{Cl}^-]$$

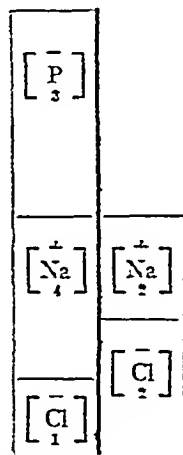
Since the anion P^- is non-diffusible it follows that the diffusible ions Na^+ and Cl^- must be unequally distributed on the two sides of the membrane. In other words, the non-diffusible ion hinders through electrostatic attraction the free diffusion of the oppositely charged Na^+ ions. The concentration of the Cl^- ion is therefore less and that of the Na^+ ion greater on the side occupied by the non-diffusible ion, thus



The ratio of the Na^+ concentrations on the two sides of the membrane (which may conveniently be styled *right* and *left*) is the same as, but reciprocal to, the ratio of the Cl^- concentrations. Thus

$$\frac{[\text{Na}^+ \text{ left}]}{[\text{Na}^+ \text{ right}]} = \frac{[\text{Cl}^- \text{ right}]}{[\text{Cl}^- \text{ left}]}$$

A difference in electric potential—*membrane potential*—is created between the solutions on the two sides of the membrane which is proportional to the logarithm of this ratio. When the ratio is 1:10 the potential difference at 19°C amounts to 58 millivolts. The ionic distribution may be represented thus



Numbers indicate concentrations in arbitrary units

The Donnan equilibrium is responsible for many physico-chemical effects of the utmost physiological importance, e.g. a difference of pH on two sides of a cell membrane across which H^+ ions can diffuse, differences of electrical potential between two sides of a membrane and differences in osmotic pressure between the interior of a cell and the extracellular fluids (see also ch. 32).

THE REGULATION OF THE ACID-BASE EQUILIBRIUM OF THE BLOOD

Acids such as phosphoric, sulphuric and hydrochloric, carbonic and lactic as well as certain organic acids are continually being formed in the processes of metabolism. In disease, e.g., diabetes,

acids such as β -hydroxybutyric and aceto-acetic acids are produced in excess. Yet in health the reaction of the blood remains remarkably constant at about pH 7.4, and even in disease may show little or no variation from the normal, for the adjusting mechanisms called into play to neutralize and remove excess acid perform their duties with extraordinary efficiency.

The reaction of the blood is protected by three lines of defense

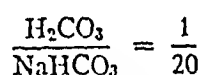
- (1) The buffer systems of the blood
- (2) Excretion of carbon dioxide by the lung
- (3) Excretion of fixed acid by the kidney

The intestinal mucosa also assists to some extent in the removal of acid, particularly of a part of the phosphoric. The role played by the kidneys in the regulation of the acid-base balance is considered in chapter 35.

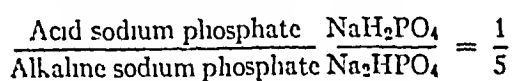
THE BUFFER SYSTEMS OF THE BLOOD may now be enumerated

I In the plasma (primary buffers)

- (1) Carbonic acid (free acid) and acid bound as sodium bicarbonate. Ratio 1 to 20

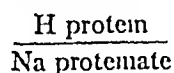


- (2) The acid and alkaline phosphates of sodium. Ratio 1 to 5



NaH_2PO_4 may be regarded as a weak acid and Na_2HPO_4 as the salt of the buffer pair

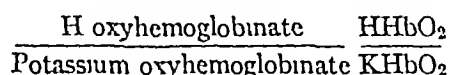
- (3) Plasma proteins which at the reaction of the blood behave as acids and so combine with base



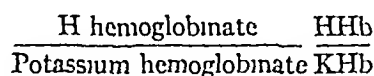
II In the corpuscles (secondary buffers) It is to be remembered that whole blood consists of two liquid phases separated from one another by the membrane of the corpuscles. Different materials in different concentrations exist within the red cell and in the plasma. The corpuscles as well as the plasma contain important buffers and an interchange of water, anions and H^+ occurs between the two across the membrane. The corpuscular membrane, on the other hand, is impermeable to K^+ and almost impermeable to Na^+ as well as to the colloidal anions, hemoglobin and plasma

protein. The buffers of the red cells are

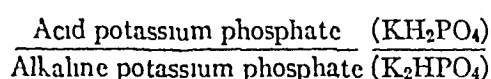
(1) Oxyhemoglobin and reduced hemoglobin act each as a weak acid of a buffer pair of which the potassium salt of the pigment acts as the other half. Thus



and

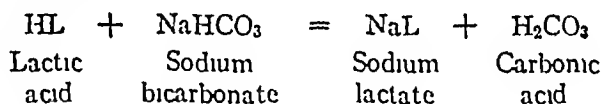


- (2) The potassium salts of phosphoric acid



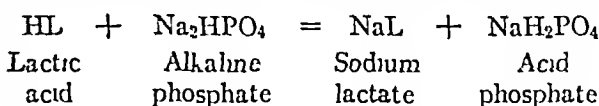
The reactions by which acids are buffered in the blood may now be considered in more detail. They may be grouped under the following heads

(1) *Fixed acids* formed during metabolism, e.g., lactic, sulphuric, phosphoric, etc., are buffered by the bicarbonate in a way analogous to that already described on page 129. Taking lactic acid (HL) as a type the reaction may be expressed by the following equation



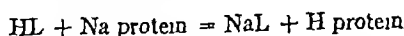
The comparatively strong fixed acid is thus replaced by a neutral salt. Phosphates, sulphates, lactates, etc., are excreted in the urine, lactic acid is also removed to a large extent through its conversion to glycogen in the liver and muscles. Carbonic acid is removed through the diffusion of CO_2 into the alveolar air. The great value of bicarbonate in maintaining neutrality of the body fluids is not due so much to its true buffering action, i.e., to the replacement of a strong by a weak acid, but to the fact that the latter is volatile and can be eliminated by the lungs.

Some of the fixed acid also reacts with the alkaline phosphate with the production of a salt of the acid and a greater proportion of acid phosphate



The ratio $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ is readjusted by the excretion of the excess acid phosphate in the urine. The phosphate mechanism for removal of acid from the body is much less efficient, however,

than in the bicarbonate system, for the reasons that (a) the excretion by the kidney is relatively slow, and (b) base is lost from the body in combination with the acid. Plasma protein also yields base to neutralize fixed acids



(2) *Carbonic acid* which as a product of tissue activity is produced in much greater quantities (800 to 900 grams daily) than any other acid, is buffered mainly by base released from hemoglobin. When carbon dioxide enters the blood the sodium bicarbonate of the plasma increases. The increase occurs to a much less extent if plasma separated from the red cells is exposed in the same way to a high tension of CO_2 . The rise in the sodium bicarbonate of whole blood is the result of interchanges between corpuscles and plasma.

It was thought at one time (Zunz) that the rise in plasma bicarbonate was due to the passage of base from the corpuscles into the plasma. It was shown by Gürber and by Hamberger, however, that the rise of bicarbonate was accompanied by a reduction in chloride. It was concluded that chloride derived from the NaCl of the plasma diffused into the red cell and the base (Na) thereby released combined with carbonic acid to form bicarbonate. This mechanism is referred to as the "*chloride shift*". The interchanges between the corpuscles and plasma are believed to occur as follows. As the blood passes from the arterial to the venous side CO_2 is absorbed and diffuses into the red cell. Herein, through the action of an enzyme—*carbonic anhydrase*—(ch 32), it is converted to carbonic acid (H_2CO_3). The oxyhemoglobin becomes at the same time reduced. Reduced hemoglobin is a much weaker acid than oxyhemoglobin and so gives up its alkali to the carbonic acid. Bicarbonate (KHCO_3) and HHb are formed. The base which previously had been bound to the non-diffusible hemoglobin is now bound as bicarbonate. The concentration of bicarbonate ions (HCO_3^-) in the cells is thereby raised above that in the plasma and, as a result, this anion diffuses across the corpuscular membrane. Since the cations cannot diffuse, the ionic equilibrium between the plasma and the interior of the cell will tend to become disturbed. The equilibrium (p 130) is sustained by the diffusion of Cl ions from the plasma into the corpuscle where they combine with base. The HCO_3^- ions which leave the cells combine with the sodium released from chloride to form plasma bicarbonate (see fig 32 8).

The exchanges just described increase the concentration of osmotically active substances within the cells, water therefore passes from plasma to corpuscle, which as a result increases in volume. Though described in steps, the reactions actually occur simultaneously. In the lungs reverse ionic interchanges occur between the cells and plasma. H_2CO_3 within the corpuscles is acted upon reversibly by carbonic anhydrase, CO_2 enters the plasma and diffuses into the alveolar air, oxygen is absorbed and diffuses into the cell. The oxyhemoglobin so formed, being a stronger acid, recaptures the base which it had lost. Cl ions leave the cells and recombine with plasma base, while HCO_3^- ions enter the cells and recombine with base released from chloride. Thus the plasma bicarbonate is reduced and the plasma chloride increased. Water leaves the cells for the plasma.

Carbonic acid is also buffered to a minor extent in the plasma through the protein and phosphate buffer systems. The plasma protein yields base and more free (acid) protein is produced. The alkaline phosphate is converted to the acid phosphate with a consequent fall in the $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ ratio.

It is evident from the foregoing description of the interchanges between the corpuscles and the plasma that whole blood as compared with plasma possesses a much greater power to buffer CO_2 . For this reason plasma separated from the red cells without precautions to prevent the escape of CO_2 is called "separated plasma" to distinguish it from plasma of whole blood, i.e., plasma in contact with the red cells. Only when precautions have been taken to prevent the escape of CO_2 during the separation of plasma from the cells will the separated plasma have the same reaction and amount of bicarbonate as existed when it was in the body. Only then will the equilibrium between cells and plasma be maintained unchanged. Such plasma is therefore referred to as "true plasma" (see also p 397).

ALKALI RESERVE—ACIDOSIS AND ALKALOSIS

THE ALKALI RESERVE

This term was brought into use by Van Slyke and Cullen to denote the amount of base in the blood which is available for the neutralization of fixed acids, e.g., lactic, hydrochloric, etc. When CO_2 enters the blood, base is liberated in the manner already described and bicarbonate, NaHCO_3 , is formed in the plasma. The plasma bicarbonate then is a measure of the base left over after all acids

stronger than H_2CO_3 have been neutralized and indicates the reserve of alkali readily available for the neutralization of such acids. The quantity of plasma bicarbonate therefore gives indirectly a measure of the extent of the production of fixed acids in the body. For if acid production be increased the bicarbonate becomes reduced, its base being given up for the neutralization of the stronger acids. The term alkali reserve, it should be emphasized, refers only to base bound as bicarbonate and not to the total base of the blood. A large quantity of base, Na, K, Mg and Ca is already bound as salts of fixed acids, chiefly sodium chloride, which the weak acid H_2CO_3 is unable to displace (fig 13 1). Though changes in the alkali reserve may result from alterations in the body's store of total base, they also occur quite independently of any such change, i.e., simply from variations in the distribution of base between carbonic acid and fixed acids.

It is also important to remember that the term alkali reserve refers to the *absolute* quantity of bound CO_2 in the plasma and has no reference to the ratio between this value and the quantity of the free CO_2 . Upon this ratio— $\text{H}_2\text{CO}_3/\text{NaHCO}_3$ —the pH of the plasma depends. The plasma bicarbonate may therefore be greatly depleted, yet if the free carbonic acid be reduced to a corresponding extent and the normal ratio of 1:20 thereby maintained, the hydrogen ion concentration will show no appreciable change. Large quantities of acid may be formed in the body yet such is the store of bicarbonate that these, as in the case of the example given on page 129, are taken care of and not until extreme depletion of the bicarbonate buffer occurs does any great change in blood reaction result.

Rapid adjustments of the ratio $\text{H}_2\text{CO}_3/\text{NaHCO}_3$ are brought about through the exquisite sensitivity of the respiratory mechanism to changes in pH, and to the chloride shift between plasma and red cells. The slightest reduction in plasma bicarbonate through its decomposition by acid is met by increased pulmonary ventilation and an equivalent reduction in the numerator of the above expression, i.e., by the elimination of CO_2 by the lungs. Low arterial and alveolar CO_2 tensions are therefore associated with a low alkali reserve.

On the other hand, when excess CO_2 is contained in the blood a compensatory increase in bicarbonate results (chloride shift). So high arterial and alveolar CO_2 tensions are associated with a high alkali reserve.

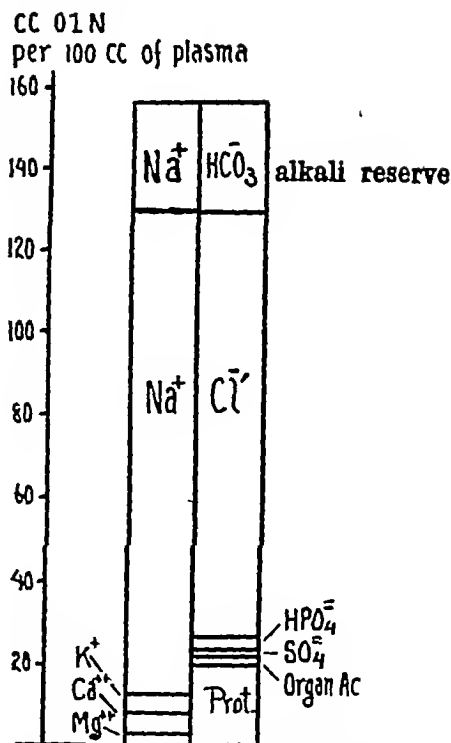


FIG 13 1 Diagram illustrating normal acid-base balance (modified from Gamble, Ross and Tisdall)

When excessive amounts of CO_2 are removed from the body (e.g., by forced breathing) a movement of chloride and bicarbonate ions occurs in the reverse direction, the plasma bicarbonate is reduced.

Measurement of the alkali reserve

The alkali reserve is measured by exposing a sample of plasma to an atmosphere of carbon dioxide under conditions similar to those existing in the body, and determining the volumes of gas which are combined as bicarbonate. This is spoken of as the *carbon dioxide combining power*, or the *carbon dioxide capacity* of the plasma. The determination is made as follows. Blood (15 cc) is drawn from a vein into a syringe containing oxalate. Venous stasis, which would cause the accumulation of carbon dioxide and so increase the plasma bicarbonate through the chloride shift mechanism (p 132) is avoided. The blood is then centrifuged under a layer of paraffin oil which prevents the escape of carbon dioxide, and so a reduction in plasma bicarbonate. After centrifuging the plasma is pipetted off, placed in a saturating vessel and equilibrated with air containing 5.5 per cent carbon dioxide, i.e., the percentage of carbon dioxide in normal alveolar air. The observer's own alveolar air is usually employed. He expires as deeply as possible after an ordinary inspiration into the saturator, through a bottle filled with glass beads, these condense the moisture from the breath, which would otherwise dilute the plasma sample. The saturator is turned end over end for a couple of minutes, the plasma is thus

spread over the interior of the vessel in a thin film. After equilibration a measured amount of the plasma (3 cc.) is introduced into a Van Slyke gas apparatus (fig 13.2). The CO_2 is then liberated by the action of an acid in the presence of a vacuum, and measured. The result gives the total carbon dioxide content of the plasma. In order to obtain the quantity combined as bicarbonate, the value for the dissolved carbon dioxide, obtained by calculation from the absorption coefficient of carbon dioxide, and the tension of the gas (40 mm Hg) at which the sample was equilibrated (ch 32), is subtracted (p 137). The normal values for the carbon dioxide combining power of plasma as so determined range from 53 to 75 volumes per cent. A reduction below 50 volumes or so would therefore constitute "acidosis", above 75 volumes, "alkalosis" (see below).

THE NORMAL ACID-BASE BALANCE AND ITS VARIATIONS

The pH of the arterial blood in health varies between 7.39 and 7.44. The normal value for the plasma bicarbonate of venous blood ranges between 53 and 75 volumes CO_2 per cent. The free CO_2 , being one-twentieth of this, amounts therefore to about 2.5 to 3.5 volumes per cent.

In disease the pH of the blood, unless in the

terminal stage, e.g., diabetic coma, never becomes actually acid, i.e., reaches a pH below 7.0, and as just stated, a reduction in the alkali reserve may occur with little or no change in blood reaction. The limits of pH compatible with life are probably not higher or lower respectively than 7.8 and 6.8. A pH of the latter value has been observed in a case of diabetic coma which ultimately recovered under insulin treatment. The pH of venous blood during rest is lower by 0.02 than that of arterial blood. The red cells are more acid than the plasma by 0.08 to 0.14 pH.

ACIDOSIS AND ALKALOSIS—DEFINITIONS

Few terms in physiology have caused more confusion than these. "Acidosis" has been used with at least two different meanings. The term was originally introduced by Naunyn to denote the production in the body of the abnormal acid metabolites, β -hydroxybutyric and acetoacetic acids.

The term was used later by Van Slyke and Cullen to mean simply a decrease in the alkali reserve (plasma bicarbonate) below the normal level. This is the sense in which the term has been usually employed. Since the bicarbonate repre-

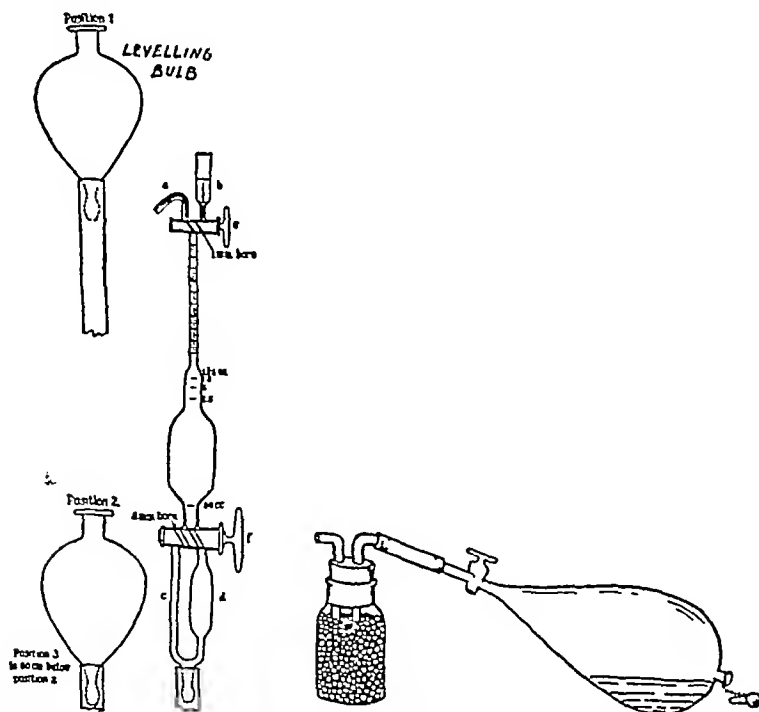


FIG 13.2 Left hand figure (Van Slyke), volumetric blood gas apparatus, right hand figure apparatus for equilibrating plasma with alveolar air (modified from Peters and Van Slyke)

sents the base in the blood which is left over after the non-volatile acids have been neutralized, a reduction in alkali reserve indicates frequently—though not necessarily—that an excess of fixed acids are being produced or retained in the body. We have seen how very efficient the compensatory mechanisms are in keeping the pH constant, and that an acidosis either in the sense of a reduction in alkali reserve or of an increased acid production may or may not be associated with a depression of the pH (reduced alkalinity). We also know that the term acidosis cannot mean any *real* acidity of the blood—a pH below 7.00—for this is incompatible with life, and only occurs when the alkali reserve is very greatly reduced and the bicarbonate buffer ratio cannot be maintained anywhere near the normal value. *Alkalosis* carries the corresponding meaning of an increased alkali reserve. This may or may not be associated with a rise in blood pH (i.e., increased alkalinity).

Van Slyke in a later paper recognizes nine possible acid-base states, one normal, and eight abnormal (fig 13.3). He avoids the terms acidosis and alkalosis, using the more precise designations, alkali deficit or CO_2 excess, and alkali excess or CO_2 deficit, respectively. In any one of these, the ratio $\text{H}_2\text{CO}_3/\text{NaHCO}_3$ may be either increased or diminished, with a consequent elevation or depres-

sion of pH, the acid-base condition is then said to be *uncompensated*. On the other hand, when adjustments occur to maintain the ratio at the normal value the condition is said to be *compensated*.

The nine acid-base states are described as follows:

(a) *Uncompensated alkali excess* In this condition the $[\text{NaHCO}_3]$ is increased without a proportionate rise in $[\text{H}_2\text{CO}_3]$. The pH is raised (i.e., the alkalinity of the blood is increased). This may result from the ingestion of large quantities of alkali (sodium bicarbonate) or from the loss of HCl by vomiting as in pyloric obstruction (area 1, in figure 13.3).

(b) *Uncompensated CO_2 deficit* In this the $[\text{H}_2\text{CO}_3]$ is reduced below normal, the $[\text{NaHCO}_3]$ is also lowered, but not sufficiently to maintain $\text{H}_2\text{CO}_3/\text{NaHCO}_3$ ratio at its normal value. The pH therefore is raised. A disturbance of this nature may result from forced breathing, or the hyperventilation induced by hot baths, high altitudes or shallow breathing (p. 426). It is sometimes also referred to as "gaseous alkalosis". Partial compensation occurs as evidenced by the reduced excretion of acid and ammonia in the urine and the increase in the urinary bicarbonate (areas 2 and 3).

(c) *Compensated alkali excess* The $[\text{NaHCO}_3]$ is raised but a parallel rise occurs in $[\text{H}_2\text{CO}_3]$. The pH is therefore normal. A disturbance of this nature will result from conditions which cause (1) but of less severe degree (area 4).

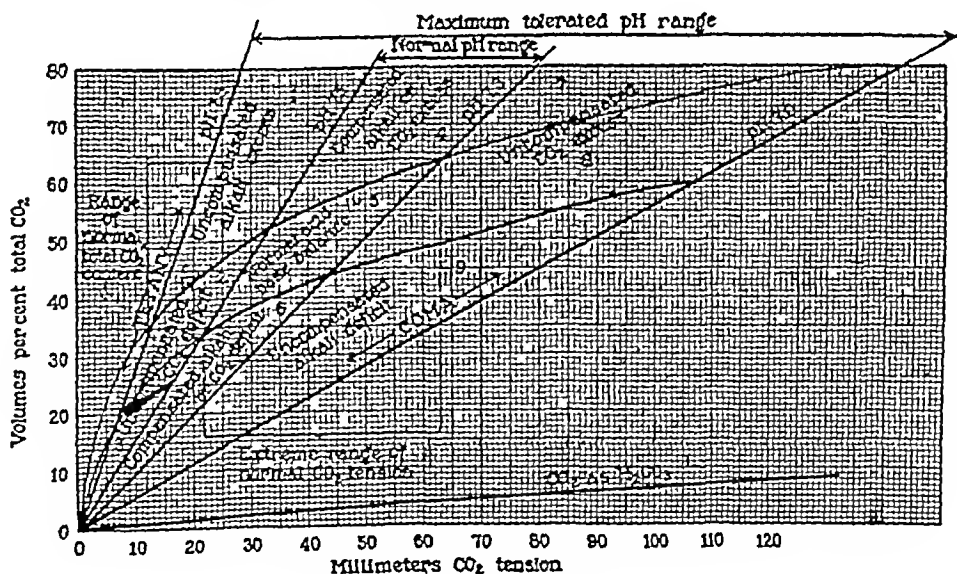


FIG. 13.3 Normal and abnormal variation of the $[\text{BHCO}_3]$, $[\text{H}_2\text{CO}_3]$, CO_2 tensions and pH in oxygenated human whole blood drawn from resting subjects at sea-level.

The curved lines are the CO_2 dissociation curves (p. 397) of reduced (upper curve) and oxygenated blood (lower curve). The straight slanting lines indicate different $\text{NaHCO}_3/\text{H}_2\text{CO}_3$ ratios, the pH at each ratio having been calculated from the Henderson-Hasselbalch equation (p. 129). The CO_2 present as bicarbonate at any point is obtained by subtracting from the total CO_2 the relatively small amount present as H_2CO_3 , indicated by the line near the bottom of the chart. The ratio, and so the pH, is the same at all points along a given line. Thus, the intersection of the CO_2 dissociation curves by these so-called isohydronic lines marks off nine areas corresponding to the acid-base states described above (after Van Slyke).

(d) *Compensated CO_2 excess* $[\text{H}_2\text{CO}_3]$ increased but accompanied by a proportional rise in $[\text{NaHCO}_3]$, pH is normal. This state occurs in conditions in which there is retention of CO_2 , e.g., emphysema (area 4)

(e) *Normal acid-base balance* $[\text{H}_2\text{CO}_3]$, $[\text{NaHCO}_3]$ and pH are normal (area 5)

(f) *Compensated alkali deficit* $[\text{NaHCO}_3]$ is reduced and $[\text{H}_2\text{CO}_3]$ diminished proportionately, pH normal. The condition results from the abnormal production of fixed acid as in diabetes, acid retention as in nephritis, or from the ingestion of mineral acids or acid producing salts (CaCl_2 , NH_4Cl). This and (9) are the acid base states to which the term "acidosis" has been usually applied. The compensatory adjustments are evident in the increased acid and ammonia excretion in the urine and the lowered CO_2 tension in the alveolar air. Alkali deficit may also result from the increased excretion of base as occurs in dehydrated states (p. 23) and in animals following the production of a pancreatic fistula (area 6)

(g) *Compensated CO_2 deficit* This is a reduction in $[\text{H}_2\text{CO}_3]$ and a parallel reduction in $[\text{NaHCO}_3]$. The pH is therefore normal. This state results from less severe grades of the conditions which cause (2) above. The excretion of NaHCO_3 by the kidney is much slower than the excretion of CO_2 by the lungs, so a condition of CO_2 deficit, which is uncompensated to start with, later tends to become compensated (area 6)

(h) *Uncompensated CO_2 excess* $[\text{H}_2\text{CO}_3]$ is increased but is not balanced by a proportional rise in $[\text{NaHCO}_3]$. The pH is lowered (i.e., blood is less alkaline). This is sometimes referred to as "gaseous acidosis". Such a state results from a hindrance to the excretion of CO_2 as may occur in pneumonia, obstruction to breathing or depression of the respiratory center by morphine. There is partial compensation, i.e., the bicarbonate is raised above the normal and there is an increased

excretion of acid and ammonia by the kidneys (areas 7 and 8)

(i) *Uncompensated alkali deficit* $[\text{NaHCO}_3]$ is reduced without there being a parallel reduction in $[\text{H}_2\text{CO}_3]$. The pH is lowered (i.e., the blood becomes less alkaline). Such an acid base state occurs when large quantities of fixed acids— β -hydroxybutyric and aceto-acetic—are produced as in diabetic coma and in the terminal stages of nephritis when acid excretion is greatly impaired. The plasma bicarbonate is severely depleted, the CO_2 combining power of the plasma may be less than 20 volumes per cent (area 9)

It will be evident after a consideration of some of the foregoing acid-base states that the terms "acidosis" and "alkalosis" if used to denote respectively a fall or a rise in plasma bicarbonate are misleading, since the blood may be no more acid or alkaline than normal. For example, in (b) above, the blowing off of CO_2 will result in a certain degree of compensatory reduction of NaHCO_3 . The reduction in the alkali reserve would therefore entitle it to be called acidosis though the blood was actually more alkaline than normal. In (h) on the other hand the blood is less alkaline than normal, yet a certain degree of compensatory increase in the alkali reserve would have occurred. The term alkalosis would therefore apply, though its use would give an erroneous idea of the true state of the acid base balance. On account of the ambiguity of these terms and the confusion which is likely to arise from their use it has been recommended by the British Medical Research Council that the term *acidemia* be used for a state in which the pH is lowered and that the term *acidosis* be restricted to indicate a lowered alkali reserve with unaltered pH. The term *alkalosis* would be retained as the corresponding term to indicate the reverse condition, namely an increase in the reserve alkalinity. The term *alkalemia* would indicate a blood state in which the pH was raised.

MEASUREMENT OF THE HYDROGEN ION CONCENTRATION OF THE BLOOD

(1) *Electrometric method*

(a) *By means of hydrogen electrodes* Two hydrogen electrodes are employed. The method is derived from the fact that in a suitable concentration cell a difference of potential between a metal electrode (e.g., mercury) and a solution of its ions (e.g., mercurous chloride) is set up which is proportional to the ion concentration. A hydrogen electrode, composed of spongy platinum saturated with hydrogen gas, is therefore used to determine the hydrogen ion concentration of a solution

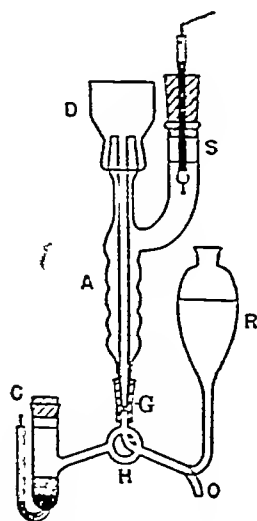


FIG. 134. Description in text.

The platinum is contained in a small tube and partially immersed in a normal solution of KCl. The tube above the solution is filled with hydrogen gas, which is adsorbed by the platinum. The hydrogen electrode is balanced against a calomel electrode (metallic mercury in contact with a normal KCl solution saturated with HgCl), or against another hydrogen electrode dipping into a solution of known hydrogen ion concentration. When, as in the latter instance, the two hydrogen electrodes are connected, an electromotive force is developed, which varies with the difference in hydrogen ion concentration of the two solutions. This can be measured by a suitable instrument. Another electrometric method involves the use of quinhydrone—*quinhydrone electrode*. One platinum or gold electrode is placed in a solution of known hydrogen ion concentration, and the other dips into the unknown solution. A small quantity of quinhydrone is added to each solution. The quinhydrone undergoes oxidation to quinone or reduction to hydroquinone according to the hydrogen ion concentration of the solution. A potential difference is set up between the two electrodes and from its magnitude the H ion concentration is calculated.

(b) *By means of the glass electrode* Like the hydrogen electrode the potential of the glass electrode, in an electrical system similar to that just described for the latter, varies with changes in pH. It must be calibrated with solutions of known hydrogen ion concentration.

The glass electrode, as shown in fig. 13.4, consists of a tube G of thin (0.001 mm), soft glass of special composition, closed at its lower end and sealed into tube A of ordinary glass. Filling the larger tube and surrounding the surface of special glass is a 0.1 N solution of HCl, into which dips a silver-silver chloride electrode S. The solution, whose hydrogen ion concentration is to be determined, is introduced into the cup D, which fits the tube G. C is a calomel electrode and R a reservoir of saturated solution of KCl. Through the stopcock H contact is made with the glass tube containing the solution to be tested. The theory of the electrode is complicated and not fully understood. Electrolytic dissociation must occur within the special glass itself, since a minute electrical current can pass through it.

The glass electrode has largely superseded other types of electrode which have been used for the hydrogen ion determinations. It has the advantages of being less cumbersome, more durable,

being unaffected by oxidizing agents and of requiring only a small amount of fluid for testing, for the tube D need contain only a drop or so.

(2) Colorimetric method

In the method of Levy, Rowntree and Marriott as modified by Dale and Evans, the blood is placed without loss of CO₂ in a collodion sac and dialyzed for 15 minutes against a solution of normal saline. A few drops of 0.02 per cent of the indicator neutral red are added to the dialysate. The latter is then placed in a comparator and its color matched with that of a phosphate mixture of known pH and containing neutral red in the same concentration as the blood dialysate. The reader is referred to Dale and Evans' article for the details of the method.

It might be thought that the H ion concentration would be reduced through the dilution of the dialyzed substance by the saline in the apparatus. But, since the pH of a buffered solution depends upon the ratio of free to bound acid and this of course will not be altered, so, moderate dilution does not effect the result. In the colorimetric method of Cullen, the pH is determined directly (i.e., without dialysis) upon plasma which is diluted 20 times, phenol red is used as indicator. In this method a correction must be made for dilution and also for protein content.

(3) From the H₂CO₃/NaHCO₃ ratio

The total CO₂ in a sample of plasma is obtained as described on page 133. The free CO₂, i.e., the CO₂ in simple solution, is obtained by calculation from the partial pressure of CO₂ with which the sample was equilibrated. The calculation is made as follows. The figure for the partial pressure of CO₂ is multiplied by the absorption coefficient of CO₂ in plasma, which at 38°C is 0.510 (p. 366). If, then, the partial pressure of CO₂ is 40 mm Hg (as in normal arterial blood), the quantity of gas dissolved in 100 cc of plasma is

$$40 \times \frac{100 \times 0.510^*}{760} = 2.68$$

The total CO₂ is, say, 56 volumes per cent. Then the combined CO₂ is 56 - 2.68 = 53.32 volumes per cent.

The pH may now be calculated from the Henderson-Hasselbalch equation (pp. 128 and 395)

$$\text{pH} = \text{pK}_1 + \log \frac{[\text{B}]\frac{1}{2}[\text{CO}_2]}{\text{H}_2\text{CO}_3}$$

The value of pK₁† for plasma is 6.10, therefore

$$\text{pH} = 6.10 + \log \frac{53.32}{2.68}$$

* $100 \times (0.510/760) = 0.0672$ is a constant factor by which the CO₂ partial pressure (in mm Hg) is multiplied to give the volumes of dissolved CO₂.

† K₁ includes the first dissociation constant of H₂CO₃ and a figure representing the dissociation of NaHCO₃ under the conditions existing in plasma.

SECTION II THE CIRCULATION OF THE BLOOD

By N B T

CHAPTER 14

THE DYNAMICS OF THE CIRCULATION

GENERAL OUTLINE OF THE VASCULAR SYSTEM

The main anatomical features of the circulatory system may be very briefly recalled. The vessels constituting the vascular bed differ widely in their calibers in different regions and on the basis of their size, structure and physiological relationships are divisible into four main groups—the *arteries*, *arterioles*, *capillaries* and *veins*. The arteries are constructed to withstand a high pressure. Their walls are thick and contain a large proportion of elastic tissue and some involuntary muscle fibers, an outer sheath of connective tissue and an inner lining of endothelial cells. The outer coat is called the *tunica adventitia*, the middle coat the *tunica media* and the inner coat the *tunica intima*. The relative proportions of muscular and elastic tissues vary with the size of the artery. The largest vessels, such as the aorta and the pulmonary artery, are relatively poor in muscular tissue but contain a large proportion of elastic fibers. The medium sized arteries contain a relatively large amount of muscle and less elastic tissue, while in the smaller vessels the muscle is greatly in excess. The walls of the arteries are supplied with minute vessels—*vasa vasorum*—which ramify in the outer and middle coats. They also are furnished with nerves. As the arterial system is traced peripherally, the vessels are found to break up into innumerable branches whose calibers become reduced with successive divisions. But with each subdivision of the arterial system the *total cross area* of the vascular tubing increases. The sum of the cross-areas of all the capillaries, for example, is many times greater than the cross-area of the aorta. This fundamental fact in the dynamics of the circulation should always be borne in mind. The final divisions of the arterial system proper are minute vessels called *arterioles*. These vary in size but on the average are about 0.2 mm in their outside diameters. Their walls are relatively thick and composed almost entirely of smooth muscle lined by an endothelial layer and sheathed by a scanty *adventitia*. The muscle fibers are supplied with excitator and inhibitor nerves (ch 27). The *arterioles* after a course of variable length lose their muscular and connective tissue coats while the inner endothelial tubes that remain are continued on as extremely fine, hairlike vessels, the *capillaries*. The capillaries, a number of which arise from a single *arteriole*, are from 2 to 18 microns in diameter, depending upon volume of blood flowing through them, and from a half

to one millimeter long. Lying upon the outer surface of the capillary wall in amphibians and certain other cold blooded vertebrates a scattering of peculiar cells with a number of long processes is to be seen. These are called after their discoverer, Rouget cells (see also ch 28). The processes of the neighboring Rouget cells join with one another to form a loose mesh work which encloses the capillary. The capillaries form maze like plexuses with one another and connect the arterial and venous systems. Their venous ends converge to form first the smaller *veins* or *venules*. By the confluence of these to form larger channels, and the successive junctions of veins of ever increasing caliber, the blood is finally poured into the right auricle by two large trunks, the superior and inferior *venae cavae* whose cross-areas are together little more than double that of the aorta.

The veins have much thinner walls than the arteries, but like these they possess three coats—*intima*, *media* and *adventitia*. The middle coat is only a fraction of the thickness of that of a corresponding artery. It is composed of a relatively small amount of unstriped muscle and a large amount of connective tissue. The elastic tissue is scanty. The outer coat of the vessel is disproportionately thick, being several times thicker than the *tunica media*. The valves which are present in some of the larger veins, such as the femoral, are formed by foldings of the *tunica intima*.

The course of the blood through the body from the left ventricle to the right auricle is spoken of as the *greater* or *systemic circulation*. From the right auricle the blood enters the right ventricle from which it is discharged into the pulmonary artery and thence through the *lesser* or *pulmonary circulation*. The heart serves as a two-cylinder pump situated between these two systems. The left ventricle receives arterial blood from the lesser circulation, and drives it through the systemic vessels. The right ventricle is supplied with venous blood from the systemic vessels and pumps it through the pulmonary circuit. Both sides of the heart contract synchronously and obviously must eject the same quantity of blood in a given time. Otherwise blood would be dammed back in some part of the circulation.

THE PRINCIPLES GOVERNING THE FLOW OF A LIQUID THROUGH A SYSTEM OF TUBES

The vascular system may with advantage be considered here in its purely physical aspect, and

its more physiological characters be left to be dealt with later. The movement of the blood within the vascular channels is governed by those physical laws which govern the flow of liquids in any other closed system of cylindrical tubes. For this reason the dynamics of the circulation may be more simply explained and the more readily understood from illustrations furnished by artificial models.

Figure 14 1 represents a model comprising a reservoir and a long horizontal tube from which upright tubes of the same bore and numbered 1 to 6 lead off. The horizontal tube can be closed or opened to any desired extent by means of the stop-cock T. The level of fluid in the reservoir is at all times kept constant by means of the faucet.

If the stop-cock is closed and water poured into the system, the fluid will rise to exactly the same heights in the reservoir and all the side tubes. The heights of the liquid in the reservoir and in the upright tubes represent a certain amount of *potential energy*—energy of position. If the fluid were allowed to flow, a certain proportion of the potential energy would be converted into *kinetic energy*—energy of motion. The total height of the column of fluid in the reservoir is spoken of as the *reservoir head*, H . The height to which the column has risen in a given side tube represents the pressure that is exerted at that point upon the wall of the horizontal tube—that is, the lateral pressure. Thus if the column, say, is 100 mm high then the lateral pressure in the horizontal tube at that point is obviously 100 mm of water. In this model the heights of the narrow columns are the same as the reservoir head, therefore the lateral pressure and the total potential energy are of equal value, in other words, the entire potential energy available is exhibited as pressure.

Let us now suppose that the stop-cock is opened to its fullest extent or removed entirely as in figure 14 2. Let it also be assumed that the outflow of liquid is at a hypothetical maximum—just as though a large opening were made in the bottom of the reservoir chamber—and that there is no friction or other impediment to the outflow. In such an instance the entire potential energy would be transformed into energy of flow (kinetic energy). The fluid would not rise at all in the side tubes, that is, there would be no lateral pressure. The fluid would leave the system at a velocity equal to that which the same body of water would attain if it fell free through a vertical distance equal to its own height, i.e., the height of the reservoir column. From the two foregoing examples of two extremes, it is seen that a given amount of potential energy may remain as such, or, theoretically, be converted entirely into kinetic energy.

In figure 14 3 the stop-cock is opened. Part of the total potential energy is now utilized in giving velocity

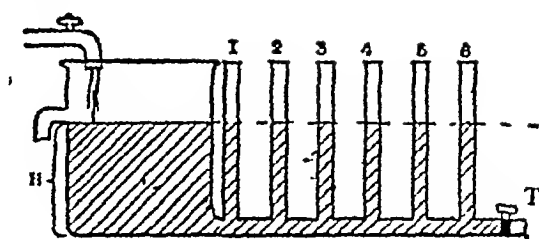


FIG 14 1 Description in text

to the flow of fluid in the horizontal tube. The remainder is exhibited as lateral pressure which can be expressed as the height in centimeters of the fluid in the side tubes. The velocity of flow (kinetic energy) will increase or diminish and the lateral pressure fall or rise reciprocally. From these considerations it may be stated that *the greater the velocity of outflow from the system the lower will be the lateral pressure and vice versa*.

When, as in figure 14 3, the stop cock is partially open and, as in practice, the flow along the horizontal tube is impeded by friction, the water rises only part way in the side tubes, and not to the same height in each. The level of the column in the first tube is a little lower than that of the fluid in the reservoir which is kept constant by the supply faucet. The second column is lower than the first, the third than the second and so on, there being a steady decline in the levels of successive tubes. A line joining the menisci is straight, and falls a little below the surface of the fluid in the reservoir.

The difference between the level of the first column and that of the reservoir is evidently due to the flow of liquid along the horizontal tube, since when, as shown in figure 14 1, there was no flow the two levels were the same. This difference between the height of the reservoir and that of the first column represents the proportion of potential energy which has been converted to kinetic energy, i.e., which is responsible for the velocity of flow along the horizontal tube. It is therefore termed the *velocity head*, V . That the drop in lateral pressure in tube 1 is actually due to the velocity of the flow may be shown by having its lower end project into the horizontal tube and bent so that its opening faces the stream as in figure 14 4. When this is done the velocity of the fluid causes the level of column 1 to attain practically the same height as the reservoir.

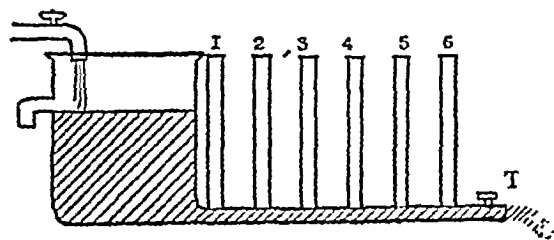


FIG 14 2 Description in text

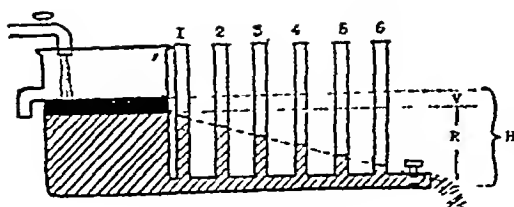


FIG 14.3 Description in text.

column. If the opening is turned in the opposite direction as in tube 3 then the level of the column is depressed an equal distance below its original level as a result of the velocity in the horizontal tube.

The difference in the water levels of successive tubes cannot, however, be explained by a progressively greater amount of potential energy being converted into kinetic energy; the velocity of flow must be the same throughout the length of the horizontal tube since its bore is uniform and in a given time the same amount of fluid that enters at one end from the reservoir must leave by the other. Furthermore, if a tube, (6) say, be extended and turned to face the stream (fig 14.4) the fluid in it rises, but the rise in this tube only partially annuls the original fall in its level and is precisely equal to the rise which occurred under the same circumstances in tube 1, i.e. it is equal to the velocity head.

The progressive fall in lateral pressure along the system is due to the energy lost in overcoming friction. At the outlet the lateral pressure has fallen to zero. Evidently then at this point the whole of the potential energy not used in giving velocity to the fluid has been consumed in overcoming frictional resistance. So, R (fig. 14.3) will represent this lost energy and is called the *resistance head*. The proportion of energy used in this way is very small in tube 1 since the length of horizontal tube traversed by the fluid is very short, but increases at the expense of the lateral pressure, in proportion to the distance along the horizontal tube traversed by the fluid and the greater frictional resistance that must therefore be overcome.

If the velocity be increased by further opening the stop-cock and allowing more fluid to flow out two effects will be caused.

(a) The lateral pressures in all the tubes will show a greater fall below the level of the reservoir column since

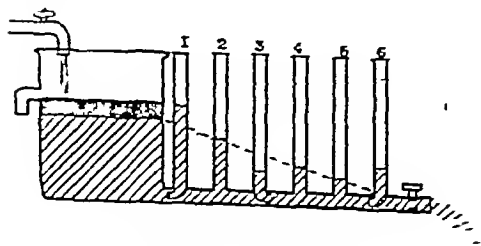


FIG 14.4 Description in text.

more potential energy will be converted into kinetic energy, the velocity head is increased.

(b) The slope of a line joining the menisci of the fluid columns will be steeper, i.e., the differences in pressure between successive tube levels—the *pressure gradient*—will be greater, more energy will be used in overcoming friction, for the resistance due to friction increases approximately in proportion to the square of the velocity (See 3 below). Reducing the outflow from the tap will produce the reverse effects.

On the other hand, if the diameter of the outlet from the horizontal tube remain the same but the height of the fluid in the reservoir be raised or lowered (thus increasing or decreasing the pressure gradient), then the velocity will vary proportionately.

Some of the more important laws governing the flow of viscous liquid through rigid tubes of small caliber may now be briefly summarized.

- (1) *Lateral pressure is inversely proportional to the velocity (Bernoulli's effect)*
- (2) *Resistance varies with the length of the tube, i.e., is proportional to the frictional surface. The pressure falls, therefore, progressively, along the tube*
- (3) *Resistance is proportional, approximately, to the square of the velocity, with a constant velocity the resistance is inversely proportional to the cross area of the tube*
- (4) *The sectional area of the tube remaining constant, then the velocity of flow of a given fluid is proportional to the pressure gradient*
- (5) *The pressure head remaining constant, then the velocity is directly proportional to the sectional area of the tube, but the quantity of fluid passing through the tube per unit of time—volume flow—is proportional to the fourth power of the radius of the tube and inversely proportional to the length of the tube (Poiseuille law)*

So far, a model has been dealt with in which the horizontal tube is of uniform caliber throughout. But if the diameter of the tube at any part be increased the velocity of flow through this part will diminish, conversely the velocity will increase through a narrowed section of the tube. We must therefore state a sixth law.

- (6) *In a tube of varying diameter the velocity varies inversely and the lateral pressure directly (see (1) above) with the sectional area of the tube*

In figure 14.5a, the horizontal tube is shown with a dilated region near its center. Below the model is a more diagrammatic representation (b). The arrows indicate the flow of liquid through the tube. Obviously, since liquids are practically incompressible, the same quantity of water must, in a given time, leave the tube as enters it. Consequently, equal quantities of water must take the same time to pass any point along the tube. Section 2 of the tube has precisely the same capacity as section 1 or 3. But the length of section 2 is only half that of either of the other sections. It is

clear then that the body of water in section 2 has only half the distance to travel as has an equal body in either section 1 or 3. The velocity of the stream must therefore be reduced by half.

Since the lateral pressure is inversely proportional to the velocity (first law, p. 140), then the level of the fluid column in tube 4 (fig. 14.5) leading from the dilated section will be higher than the levels of the tubes on either side. These facts explain why an abnormal dilatation in the vascular system, such as an aneurysm or a varicose state of the veins tends to increase. A vicious circle is established and the sacculation tends "to go from bad to worse", since the pressure is exerted over a greater area of the vascular wall. The widened bed of the stream also causes a local reduction in velocity and a consequent increase in pressure. These factors in turn cause a stretching of the vascular walls and further dilatation, and so on.

It is immaterial whether the increase in the cross area of the stream is brought about by a single dilatation or is accompanied by a division of the tube into a number of smaller channels as obtains in the vascular system (cf. fig. 14.5, c and fig. 14.9, p. 144). So long as the sum of the sectional areas of all the subdivisions is increased, no matter how small these are individually, a reduction in velocity results.

The horizontal tube of the model in figure 14.5a is narrowed between the upright tubes 1 and 3. The velocity in this section is therefore increased and the pressure reduced as is shown by the height of the fluid in the second side tube.

In the final model, figure 14.6, the tap has been placed half way along the horizontal tube so as to divide the system into two sections. If the tap be partially closed, a sharp drop in the curve of lateral pressures will occur between the tubes on the left of the tap, and those on the right. That is, the model has been divided into a high and low pressure system. Increasing the flow through the tap will, as we know, cause the pressure in the left hand section to fall and the pressure in the right hand section to rise. Reducing the flow through the tap, on the other hand, causes contrary effects in the two systems.

THE APPLICATION OF SOME OF THE FOREGOING PRINCIPLES TO THE CIRCULATION OF THE BLOOD

The heart, of course, is the source of the energy whereby the blood is driven through the system at a given pressure and velocity.

THE ENERGY OF THE VENTRICULAR CONTRACTION—THE WORK OF THE HEART

The energy derived from the contraction of the cardiac muscle is expended mainly in overcoming the frictional resistance which opposes the flow of blood through the systemic vessels from the

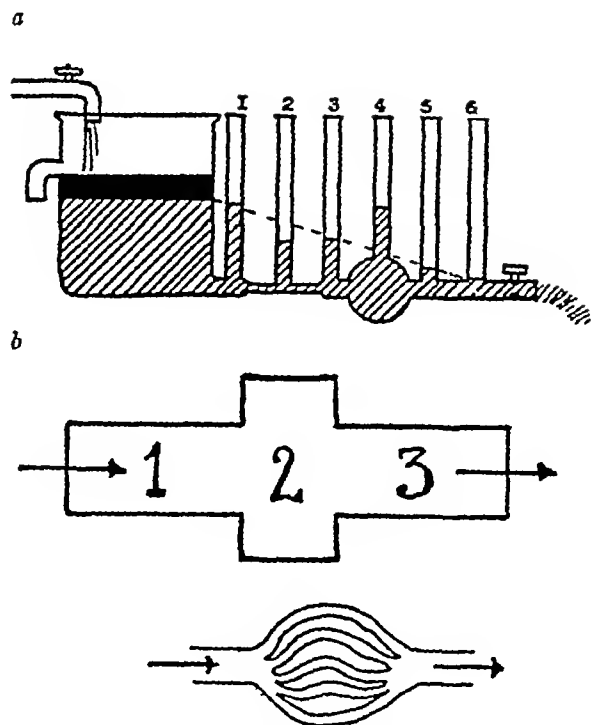


FIG. 14.5 Description in text

left to the right side of the heart in the case of the left ventricle, and through the pulmonary circuit in the case of the right. The energy expended in this way and dissipated as heat amounts to from 80 to 95 per cent of the total energy (as calculated below). The remaining fraction of the total energy liberated by the cardiac muscle appears as mechanical or external work. The latter was found by Evans and Matsuoka, employing the heart-lung preparation (p. 251), to amount to 5 or 6 per cent of the total energy when the heart was beating quietly and its output small. This figure, which represents the efficiency of the cardiac machine, is therefore very low under these conditions. The efficiency rises, however, to 20 per cent or higher when, as in muscular exercise, the output of blood from the heart per minute is maximal and the cardiac action is vigorous.¹ That is to say, the efficiency of the cardiac muscle rises up to an optimum load as its work increases. Also for a given load it works more efficiently at slow than at rapid rates.

The total energy expenditure of the cardiac muscle can be calculated from the oxygen consumption of a

¹ The difference in efficiency under the two conditions appears to be due simply to the fact that with light work the constant amount of energy expended, i.e., the oxygen consumption required to maintain the basal metabolic processes, is a greater proportion of the total energy than when heavy work is performed.

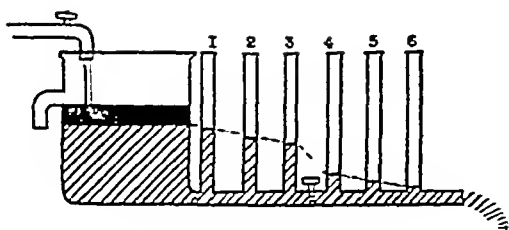


FIG 14-6 Description in text.

heart lung preparation (p. 251) Now, in the consumption of 1 liter of oxygen the heat which is generated amounts to about 5 large calories. The work equivalent of one Calorie is 426.5 kilogram meters (or 3086 foot-pounds). The consumption of 1 liter of oxygen, therefore represents a total energy expenditure of $426.5 \times 5 = 2132.5$ kilogram meters. The proportion of the total energy expenditure appearing as work gives the efficiency of the cardiac machine. The actual work performed is calculated from the following formula—

$$W = \frac{7}{6} QR + \frac{mV^2}{g}$$

Where W = total work in kilogram meters per hour, Q = quantity of blood expelled in kilograms per hour, R = mean arterial blood pressure in meters of blood. This latter is obtained by multiplying the figure for the arterial blood pressure in millimeters of mercury by 0.135; mercury is about 13.5 times heavier than blood. The pressure in the right ventricle is only one sixth of that developed in the left, so to obtain the value of QR for the whole heart this factor is multiplied by $\frac{1}{6}$.

The factor QR means, in brief, that a certain amount of blood Q is expelled from the heart against a resistance R and that the work done is equivalent to that which would be required to raise the weight of blood through a certain height. This latter corresponds in a given instance as mentioned above to the arterial blood pressure (ch. 15). When the output of the heart is small this part of the equation represents by far the greater part of the work performed. When the body is at rest, 99 per cent of the work performed by the heart (which as mentioned above is from 5 to 6 per cent of its total energy expenditure) is employed in raising and maintaining the pressure of blood in the arterial system and in expanding the elastic arterial walls.

The factor mV^2/g represents that portion of the work performed by the heart in giving velocity to the blood. That is, it gives the amount of total energy that is converted into kinetic energy. When the cardiac output is small this factor as just indicated is negligible, amounting to less than 1 per cent of the total work performed.

In the above expression m = weight of blood in kilograms per hour, V = mean velocity of the blood in meters per second at the root of the aorta, g = accelera-

tion due to gravity, a constant amounting to 9.80 meters per second, per second.

It has been pointed out by Evans that since in calculating the kinetic energy of the cardiac contraction it is the mean velocity of the blood as it is expelled from the ventricle, i.e., during the ejection period of the cardiac cycle, that is required rather than the mean velocity of the blood in the aorta throughout the entire cardiac cycle, a correction must be employed to obtain the true value. The velocity of the blood issuing from the ventricle rises to a maximum about the middle of, and falls to zero at the end of the ejection period. The velocity varies inversely with the length of this period. Therefore, when the latter has the usual duration of about $\frac{1}{3}$ of the cardiac cycle, the value for the velocity of the blood in the aorta must be multiplied by the square of $\frac{1}{3}$ or by $\frac{1}{9}$, approximately. The equation therefore becomes $\frac{7mV^2}{g}$, or the more generally applicable one, $\frac{m(VC)^2}{gE^2}$, in which C = duration of cardiac cycle and E = duration of ejection period.

The kinetic factor has the same value for each ventricle, since identical quantities of blood are ejected at equal velocities from each side of the heart at each beat. When the output of the heart is large the kinetic factor can be no longer neglected in calculating the work of the heart. It reaches an importance comparable with the pressure factor, QR . The magnitudes of these two factors may be compared by the following numerical examples under conditions of small and large systolic outputs, respectively.

A Heart of dog, small output, 6 kilograms per hour

Mean arterial blood pressure 100 mm Hg. Mean velocity of blood at the root of the aorta, 0.085 meter per second.

$$\begin{aligned} W &= \frac{7}{6} QR + \frac{7mV^2}{g} \\ &= \frac{7(6 \times 100 \times 0.013)}{6} + \frac{7 \times 6(0.085)^2}{9.8} \\ &= 9.10 + 0.031 \\ &= 9.131 \text{ kilogram meters per hour} \end{aligned}$$

B. Forcefully beating heart, with output of 90 kilograms per hour

Mean arterial blood pressure 100 mm Hg. Mean velocity in aorta 1.27 meters per second.

$$\begin{aligned} W &= \frac{7}{6} QR + \frac{7mV^2}{g} \\ &= \frac{7(90 \times 100 \times 0.013)}{6} + \frac{7 \times 90 \times (1.27)^2}{9.8} \\ &= 136.5 + 103.7 \\ &= 240.2 \text{ kilogram meters per hour} \end{aligned}$$

The kinetic factor in the above example is seen to con-

stitute a large part of the total work performed. If the output is still further increased, say to 120 liters per hour, then the value of this factor may exceed that of the pressure factor QR , it may constitute 60 per cent of the total work. The disproportionate increase of the kinetic factor over the QR factor is due to the fact that the latter increases in direct proportion to the output, whereas the former increases as the cube of the output. For instance, when as in the above examples the output is increased from 6 to 90 liters per hour, that is 15 times, the QR factor is increased to the same extent ($9 \cdot 10 \times 15 = 136$). The value of the kinetic factor, however, is increased over 3000 times since the mass of blood is increased 15 times, and $(V)^2$ is therefore increased $15 \times 15 = 225$ times, and $7m(V)^2/g$ is increased $15 \times 15 \times 15 = 3375$ times. Thus the value of the kinetic factor in example A is 0.031, in example B with an output 15 times greater it is approximately $0.031 \times 15^3 = 103.7$. If the output is increased from one of very small magnitude, say of two or three liters per hour, to one very much greater, 120 liters or so, the value of the kinetic factor in the latter instance may be increased 6000 times or more.

By employing data obtained through indirect measurements of the cardiac output (ch. 26) and the arterial blood pressure (p. 152) it is possible to arrive at an approximation of the work performed by the human heart. The average quantity of blood discharged from each ventricle (Q) when the body is at rest is around 60 grams per second. If the mean blood pressure is say 100 mm Hg then the value of R is 1.35 meters. Therefore, $\frac{1}{2} \times \frac{100}{1000} \times 1.35 = 0.093$ kilogram-meters per second, or over 5 kilogram-meters per minute, is the work performed by the average heart during rest. With this small cardiac output, the kinetic factor may be neglected. Evans took the diameters of the aortic and pulmonary orifices to be 2.5 cm, the duration of the systolic discharge as 0.3 second and the ventricular output as 60 grams per beat, obtaining a figure of 0.4 meters per second as the velocity of discharge during rest.

When, on the other hand, the output is of large magnitude (over 25 liters per minute), as in strenuous exercise, the velocity of the blood as it leaves the ventricle is over 2 meters per second. The kinetic factor as calculated from the output and velocity of ejection amounts to over 10 per cent of the total work (W) performed. The latter may reach a value exceeding 70 kilogram-meters—sufficient to lift the weight of an average sized man some three and a half feet per minute.

In aortic stenosis, with marked narrowing of the aortic ring, the velocity of ejection is greatly increased over the normal even during rest. The work of the left ventricle may be nearly doubled, and the increase is due to the enormous increase in the kinetic factor. It may, even in the absence of muscular exertion, almost equal the pressure (QR) factor. That is, the kinetic fac-

tor will then increase the work of the heart by nearly 100 per cent.

In hypertension (p. 156), R , and consequently the work of the heart, is of course increased.

THE PRESSURES AND VELOCITIES OF THE BLOOD IN DIFFERENT PARTS OF THE VASCULAR SYSTEM

The arteries constitute a high pressure system which is separated from the venous or low pressure system by the *arterioles*. These latter represent the stopcocks in the artificial models (pp. 139–142). The sphincter-like arrangement of the smooth muscle in the walls of these vessels enables their calibers to be increased or diminished, when more blood passes from the arterial system into the capillaries and veins, the pressure in the arteries falls whereas that in the capillaries and veins tends to rise. If the arteriolar calibers diminish, capillary and venous pressure changes of a reverse order result.

The pressure throughout the arterial system falls in a very gradual slope from the larger to the smaller branches as a result of the greater distance over which the blood has travelled before reaching the latter vessels and, as a consequence, of a proportion of the energy derived from the cardiac contraction having been consumed in overcoming frictional resistance. The pressure fall from the aorta to the smallest arteries is 20 mm Hg or less. The lateral pressure in an artery the size of the radial is very slightly lower than that in the aorta. The greatest fall in pressure occurs beyond the small arteries, namely during the passage of the blood through the arterioles. The pressure drop in this part of the circulation amounts to from 50 to 60 mm Hg (fig. 14.7 and fig. 15.1, p. 147).

Let us now consider the causes of this abrupt fall. We find that when the blood reaches the region of the arterioles the frictional surface is greatly increased as a result of the breaking up of the vascular bed into smaller channels. Yet the total cross area of the blood stream at this point is not very greatly enlarged. The distinction between these two statements may be clarified by a reference to the diagram (fig. 14.8). The large circle A represents the circumference of the aorta. The column of blood within has a certain sectional area and a certain surface with which it and the arterial wall are in contact. Frictional resistance is proportional to the extent of this surface. In B, which represents the arteriolar region, a somewhat larger sectional area is composed of the sectional areas

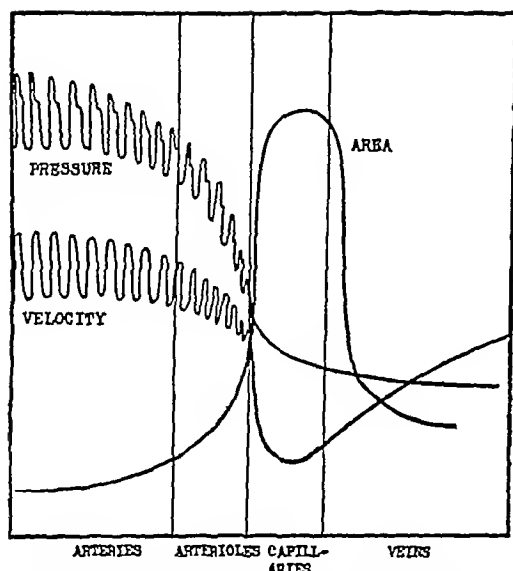


FIG 147 Diagram showing the pressure and velocity of the blood in different parts of the vascular system, i.e., from left ventricle to right auricle (modified from Fredericq) Note the relation between blood velocity and vascular area, and the absence of rhythmic variations in pressure and velocity in the capillaries and veins (see also p 149)

of a very large number of separate blood columns. The sum of all these cylindrical surfaces must be enormously greater than that of the single aortic blood column. Since the total sectional area of the vascular bed in this situation is only moderately increased, the blood velocity shows only a moderate reduction (p 140). And, since the frictional resistance is proportional to the square of the velocity, the effect of the increased surface upon the resistance to the flow of blood through the arterioles will be very great.

In the case of liquids flowing in stream lines through narrow tubes, friction is not developed between the moving liquid column and the inner surface of the tube but between the molecules of the liquid itself. The outer

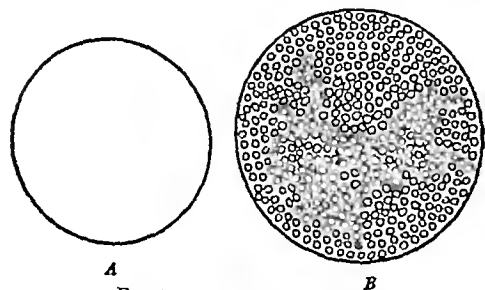


FIG 148 Description in text

layer of flowing liquid in contact with the lining of a tube is stationary. No friction therefore develops between this layer and the walls of the tube. The tubular layer of liquid next within this stationary layer rides or "slips" upon it, the third layer glides upon the second, the fourth upon the third, and so on. Thus the cross section of a moving stream consists of a series of concentric layers like the skins of an onion. For this reason simple roughening of the interior of an artery causes little increase in the resistance to the flow of blood. The movement of one layer upon the next diminishes progressively in a central direction so that in the axial portion of the stream the liquid is free from friction and its velocity is maximal. This more rapid movement in the axis of the stream can be observed in blood flowing in the small vessels of a transparent tissue. The corpuscles near the vessel wall progress comparatively slowly while those in the axial stream are swiftly moving. The greater the viscosity which the liquid possesses, the further centrally do the frictional layers extend and the smaller in extent is the frictionless "core". For a given liquid the *absolute* depth of the layers is practically the same whether the diameter of the liquid column is great or small. Consequently in the aorta where the cross area is great the proportion of this wherein friction is developed is quite small as compared with that of the slender blood columns in the arterioles where the greater part of the stream may be occupied by frictional layers.

STREAMLINED FLOW IN THE VESSELS

The blood fluid laminated as it were in the manner described flows smoothly and steadily through the vessels in stream lines. That is to say, in straight or curved lines parallel with the vascular walls, each particle of the fluid follows in the same path as the preceding one. Such a movement entails the minimum expenditure of energy. The pressure available from the heart's contraction is translated without loss into a propulsive or driving force which moves the blood through the smallest vessels. Increase in the velocity of flow above a certain critical level, or some obstruction, marked local

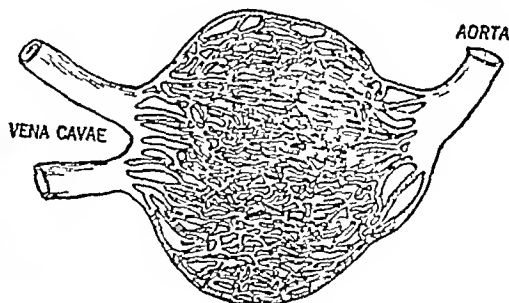


FIG 149 Diagram illustrating the relative sectional areas of the aorta, capillary bed, and great veins. The area of the "capillary lake" is much greater relatively than is shown.

ized narrowing of the vascular bed, or massing together of the corpuscles, causes eddies to be set up. The fluid particles and the corpuscles themselves move in curved paths at various angles to the line of the vessel, the flow is then said to be *turbulent*. Such a motion is non-propulsive, and causes a waste of energy, and may produce abnormal sounds or murmurs.

Beyond the arterioles through the capillaries and veins to the right side of the heart, the slope of pressure is again a gradual one. A further quota of energy is expended in propelling the blood along these vessels against resistance. The latter, though of very moderate degree, is such that by the time the blood has reached the right auricle the pressure has been reduced to almost zero—about 5 mm. of water (see also p. 167).

The *variations in velocity* throughout the vascular system are quite different from those of pressure (fig. 14.7). The velocity is high in the arteries (p. 176) but is reduced several hundred times in the capillaries and small veins. In the *capillary area*, though each vessel is extremely narrow, having, on the average, a diameter no greater than that of a red corpuscle, the sum of the sectional areas of these fine divisions of the vascular system is from 600 to 800 times greater than the cross area of the aorta. This broadening of the vascular bed into the “capillary lake”, as this area is sometimes called, causes a proportionate slowing of the blood stream (fig. 14.9). The reduced velocity here has obvious advantages since it is in this region that the interchange of gases and food materials,

and of waste products takes place between the plasma and the extravascular fluids. Since the velocity of the blood flow is so greatly reduced the frictional resistance in the capillary region is relatively small despite the narrowness of the individual channels. In consequence, any further fall in the pressure of the blood in its passage through the capillary is moderate. From the arterial end of a capillary to the venous end, the fall in pressure is not more than 15 or 20 mm. Hg. However, considering the short length of the vessel, about a millimeter, this is a relatively steep gradient. After traversing the capillary area the velocity of the blood increases again for, with the confluence of the smaller veins to form vessels of ever-increasing size but fewer in number, the total cross area of the vascular bed becomes progressively reduced. The sectional area of the great veins at their point of entrance into the right auricle is only about double that of the aorta. The slowing of the current in the capillaries and its quickening in the veins, though influenced from time to time by other more purely physiological factors to be considered later, are due in the main simply to the differences in the areas of the vascular beds in the two situations. Obviously this must be so since the vascular system is closed and, except as a temporary event when the capacity of a part of the system is suddenly increased or reduced, the same quantity of blood which leaves the left ventricle must, in a given period of time, be emptied into the right auricle.

CHAPTER 15

THE ARTERIAL BLOOD PRESSURE

THE EXPERIMENTAL MEASUREMENT OF THE BLOOD PRESSURE

The pressure in the crural (femoral) artery of the horse is sufficient to raise the blood to a height of between 8 and 9 feet. That is to say, the pressure exerted upon the walls of the artery is equivalent to the pressure exerted by a column of blood of this height. This comparatively great pressure was first demonstrated by the cleric scientist, the Rev. Stephen Hales of Teddington, England, over two hundred years ago (1733). A long glass tube was connected by means of a goose's trachea (which on account of its flexibility served in lieu of rubber tubing) to a brass cannula inserted into the animal's artery. When the blood was permitted to flow from the artery into the vertical tube it rose rapidly, but with fluctuating progress, until it reached a height of 8'3", and then oscillated above and below this level with each beat of the heart.

of the metal is raised a corresponding fraction of the distance (about 5 inches) that the blood itself would rise. The mercury being confined within a tube doubled into the form of a U, this rise is further reduced by half. The height of the pressure is indicated by the difference in levels of the two limbs of the U. Ever since Poiseuille in 1828 introduced this instrument—the *mercury manometer*—it has been customary to express the pressure of the blood in millimeters of mercury. In order to prevent coagulation of the blood, which was another serious objection to Hales' original method, the tubing connecting the cannula in the artery with the manometer is filled with a solution of sodium citrate, or other anti-coagulant fluid. Ludwig improved the method by making it self-recording. He placed a float upon the mercury column, to the float was fastened a long stiff wire bearing a writing point. By having the latter inscribe its movements upon a moving surface—a revolving drum covered with smoked paper (kymograph)—permanent tracings of the blood pressure were obtained.

THE SYSTOLIC, DIASTOLIC, MEAN AND PULSE PRESSURES

Hales in describing his experiment speaks of the blood column, after it had ceased to rise further in the tube, oscillating above and below a mean level. To quote his own words, "When it (the blood) was at its full height it would rise and fall at, and after each pulse, 2, 3 or 4 inches." In these same fluctuations taken by Ludwig's method small waves synchronous with the heart beats (fig 151) The crests of the waves which represent the maximal pressure correspond to the contraction or systole of the ventricle. The maximal pressure is consequently known as the *systolic pressure*. The troughs of the waves, i.e., the points of minimal pressure, coincide with the end of the resting phase or diastole of the cardiac cycle (ch 21). This level is called the *diastolic pressure*. The mean pressure is usually given as half the sum of the values for the systolic and diastolic pressures:

* This is not strictly accurate and a true numerical expression in millimeters of mercury of the mean pressure is not simply the average of the values of systolic and diastolic pressures, i.e., the sum of these values divided by 2 (arithmetic mean). The average pressure

difference between the diastolic and systolic pressures is the pulse pressure. This, clearly, is by the ejection of blood into the aorta systole. Its magnitude, other things being equal, will vary with the quantity of blood ejected from the heart at each beat.

The diastolic pressure represents the constant pressure to which the arterial walls are called upon to yield to overcome the resistance which the ventricular contraction must overcome to throw open the aortic valve.

It shows a steady but slight decline from the larger to the medium sized vessels. The systolic pressure shows a fall between the larger and the smaller arteries which though not great is more pronounced than that which occurs in the capillaries. On this account the two pressures tend to become more nearly equal toward the capillaries, the pulse pressure being reduced. The pulse pressure is the difference between the systolic and diastolic pressures, it may be altered by an alteration in one or other of these. A rise in the systolic or a fall in the diastolic will increase the pulse pressure, while a fall in the systolic or a rise in the diastolic will lower the pulse pressure. If both systolic and diastolic pressures rise or fall to an equal extent the pulse pressure remains unchanged. The mean pressure is increased as a result of a rise in either the diastolic pressure or of both together and will be reduced by a reduction of either or of both of these pressures.

SEVERAL FACTORS WHICH COMBINE TO MAINTAIN THE NORMAL ARTERIAL PRESSURE

There are five factors in number and though they concern points concerning the rôle which some of them play in the maintenance of the blood pressure, when touched upon more or less incidentally in various sections, it will be necessary to consider them more categorically. They are—

- (1) *The pumping action of the heart*
- (2) *The peripheral resistance*
- (3) *The quantity of blood in the arterial system*
- (4) *The viscosity of the blood*
- (5) *The elasticity of the arterial walls*

About the cardiac cycle, i.e., the true or geometric mean is somewhat lower than this, lying nearer the diastolic than the systolic pressure. If the pressure fell steadily from its systolic to its diastolic level about the cardiac cycle, and the pulse wave (see Fig. 15.1) inscribed a perfect triangle, then the arithmetic mean would be identical. The descending limb of the pulse curve, however, presents one or more secondary waves which prevent it from assuming the simple triangular form.

(1) *The pumping action of the heart* The means by which the cardiac contraction exerts its effect upon the arterial blood pressure is, obviously, through the quantity of blood which it is capable of discharging into the aorta in a unit period of time, i.e., upon the output of the heart per minute (minute volume, ch. 26). When more blood is forced into the already filled arterial system, it cannot escape at once from the system in the same amount as it is thrown into the aorta, so the arterial walls become stretched. The pressure rises until the velocity of flow through the arterioles (see below) is great enough to balance again the outflow from the system with the inflow. Hales grasped this fundamental fact when he wrote, "the real force (pressure) of the blood in the arteries depends on the proportion which the quantity of blood thrown out of the left ventricle in a given time bears to the quantity which can pass through the capillary arteries (arterioles) into the veins at that time."

(2) *The peripheral resistance* The peripheral resistance is dependent upon the caliber of the small vessels, mainly of the arterioles and, to a less extent, of the capillaries, and upon the viscosity of the blood. By far the greater part of the peripheral resistance of the circulatory system is constituted by the minute vessels of the muscles and of the abdominal structures. The importance in this connection of the latter—the so-called splanchnic area—can be demonstrated by tying off all branches of the aortic arch, except the carotids, and the abdominal aorta below the inferior mesenteric branch, when little or no change in peripheral resistance results, whereas an increase in the latter at once occurs when the blood supply to the splanchnic area is reduced by ligating the

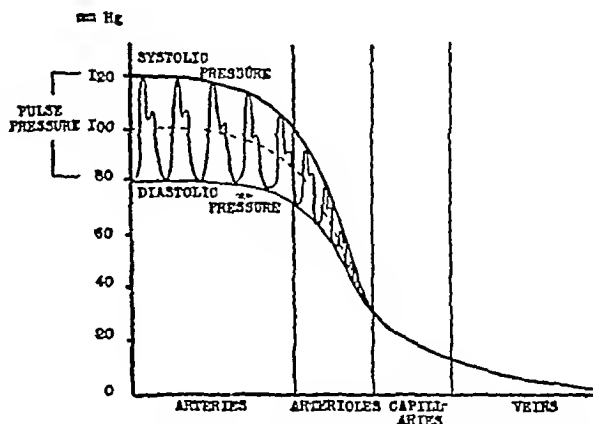


FIG 15.1 Showing the phases of the arterial blood pressure

superior mesenteric artery Stimulation of the great splanchnic nerve (ch. 27) which innervates the rings of muscle fibers in the walls of these vessels causes their constriction, and consequently a reduction in the outflow from the arterial system. As outlined in the last section the pressure will continue to rise until inflow and outflow are again balanced. In the absence of compensatory changes in the other factors concerned in the maintenance of the pressure this remains at the higher level so long as the constriction persists. Dilatation of the vessels, i.e., reduction in peripheral resistance, will of course be followed by the opposite effect. When the vessels of the splanchnic area are fully dilated they are capable of accommodating almost all the blood in the body, in such an event the blood pressure would fall to zero. The peripheral resistance might be compared to a dam in a river. If the dam is raised or lowered, and no change occurs in the supply of water flowing down the river to the dam, the water continues to rise or fall respectively (and its pressure in consequence increases or diminishes) until it reaches the new level. From then on the quantity of water which overflows in a given time is the same as it was at the original level.

The peripheral resistance varies directly with the mean arterial blood pressure and inversely with the rate of blood flow. A linear relationship within a range of moderate pressure changes exists between the ratio of these two factors, $\frac{M}{F}$ (mean arterial blood pressure)

$\frac{M}{F}$ (rate of blood flow) and the peripheral resistance. An increase or a decrease in peripheral resistance either in a separate organ or in the body as a whole (total peripheral resistance) can, therefore, be deduced from measurements of both factors, though not from one alone. For example, if the rate of blood flow increases and the mean arterial pressure falls or remains unchanged, a reduction in peripheral resistance has occurred, a rise in mean pressure accompanied by a reduced or constant blood flow indicates a rise in peripheral resistance. Green has defined a unit of peripheral resistance (PRU) as equal to $\frac{1 \text{ mm. Hg } M}{1 \text{ ml./minute}}$. It corresponds to the ohm in electricity.

The total peripheral resistance in animals or in man can be calculated from the mean blood pressure (M) and the cardiac output (F), since all blood entering the aorta must, of course, pass through the peripheral vessels.

By expressing P in dynes per square centimeter and F as milliliters per second, then the peripheral resistance can be given in absolute units of force. Thus,

$$\text{Peripheral resistance}(R) = \frac{M(\text{mm Hg}) \times 1332}{F(\text{ml./sec.})} =$$

$$R = \frac{\text{dynes/cm}^2}{\text{cm}^2/\text{sec}} =$$

$$R = \frac{\text{dyne secs}}{\text{cm}^5}$$

1332 is a figure for the conversion of pressure to dynes.

The peripheral resistance so calculated is found to vary inversely with the size of the animal, i.e., directly with the surface area. It amounts normally to from about 600 to 2,000 absolute units in man, but may be over 5,000 in arterial hypertension in which F shows little change. The values for the dog and rabbit range from 2,000 to 9,000 and from 11,500 to 12,000, respectively. This means that the minute vessels in a large animal offer less resistance than do the fewer number in a smaller animal, even though the vessels of both are constricted to the same degree. This is due to the fact that the rate of flow in the smaller animal is greater in relation to the size of its vascular bed (cardiac output bearing a constant relationship to surface area (p. 264), which varies inversely with body weight). Bazett and his colleagues use a formula into which they have introduced a correcting factor for surface area.

$$R = K \frac{M}{F/A},$$

R is peripheral resistance, K is a constant with a value of 3, and A the surface area in square meters.

(3) *The quantity of blood in the arterial system*
In any closed system of rigid tubes fluid must fill it to capacity in order that a pressure can be developed within it. The arterial walls are distensible and elastic, and a certain degree of stretching of these must occur before any considerable pressure is created. The arterial system must be actually over-filled and the greater the extent of the over-filling the greater will be the blood pressure. Loss of blood, either of all its elements, as in hemorrhage, or of the fluid portion alone, if not compensated for sufficiently by readjustment of the other factors concerned in blood pressure maintenance, must inevitably result in a fall of pressure. Increasing the total amount of circulating fluid artificially as by the transfusion of blood or blood substitute will elevate the pressure again. In animals the blood pressure may be lowered by hemorrhage to half its normal value and restored again to its original level by re-introducing

into the circulation the blood which has been removed or by the infusion of an effective blood substitute (see Effects of hemorrhage, p 27)

(4) *The viscosity of the blood* The greater the viscosity or "thickness" of any liquid the greater is the pressure required to force it along a length of narrow tube in a given time, or if the pressure remains constant the longer will be the time required for the liquid to traverse the tube. The frictional resistance which is developed between the parts of the liquid itself, that is, the internal friction (p 144) is greater when the viscosity is high than when it is low. Viscosity depends upon the degree to which the molecules or particles of a fluid cohere. Blood is some 5 times more viscous than water.³ With regard to the influence of viscosity upon the blood pressure, it is again a matter of outflow through the arterioles. If the driving force remains constant and the caliber of the vessels is unchanged, then the greater the viscosity the greater will be the frictional resistance developed in this region and the less will be the quantity of fluid that will pass through in a unit of time.

The blood owes its viscosity to its colloids (plasma proteins) and to an even greater extent its suspended corpuscles, friction is developed between the surfaces of the latter and the surrounding fluid. Changes in the concentration of the blood as a result of changes in its protein content or in the number of its corpuscles will therefore alter its viscosity, venesection by removing a quantity of blood and causing dilution of the remainder (p 17) causes a fall in viscosity which may materially relieve the work of the heart. For these reasons the viscosity is low in anemia and high in polycythemia, leukemia and anhydremia. Also changes in its chemical composition or in its gas content may alter the viscosity of the blood. Carbon dioxide increases the viscosity, oxygen lowers it, venous blood is, in consequence, more viscous than arterial, and high blood viscosity is usual in congestive heart failure with cyanosis. Chloroform anesthesia and narcotization with morphine are said to increase the blood viscosity. It is also raised in hyperglycemia, hypercalcemia and in acidosis.

The viscosity of most liquids is reduced by a rise of temperature—hot syrup flows more freely than cold. In muscular exercise and in fever the blood temperature is raised, the viscosity of the blood is lowered and the work which the heart is called upon to do in overcoming the frictional resistance in the smaller vessels is thereby appreciably reduced. Blood concentration, how-

³ This is an average figure. Values obtained by different observers for the viscosity of blood, taking distilled water as unity, vary considerably but the majority range between 4.5 and 5.5.

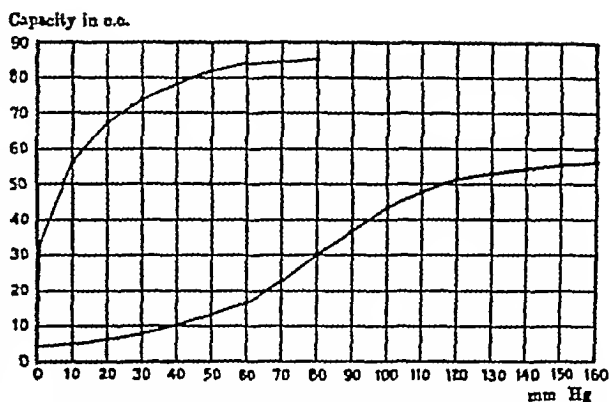


FIG 15.2 Curves of distensibility of a vein (upper curve) and of an artery (lower curve). The figures at the left side of the diagram represent the capacity of a section of the vessel when distended under a certain pressure, expressed by the figures on the base line in mm Hg (after Starling, constructed from figures given by Roy)

ever, which occurs to some extent under these circumstances, tends to offset the effect of temperature.

(5) *The elasticity of the vessel walls* This is concerned mainly with the origin and maintenance of the diastolic pressure and with sustaining the mean pressure at a higher level than would be possible in a rigid system under otherwise identical conditions.

The elasticity of arterial tissue does not come into play to any notable extent with a pressure below from 30 to 40 mm of mercury (fig 15.2). Below this level there would be little stretching of the walls of the arteries which would then behave like a system of rigid tubes. At the usual diastolic pressure that exists, however, the walls are stretched and by virtue of their elasticity tend to recoil against the distending force. We have seen that the flow of blood is pulsatile in the arteries. Beyond the arterioles, i.e., in the capillaries and veins, the flow is continuous. The conversion of the pulsatile flow to a uniform one depends upon the existence of a diastolic pressure. The physical principles involved in the maintenance of the diastolic pressure and the disappearance of the pulse beyond the arterioles may be best illustrated by a simple artificial model similar in principle to one devised by Borelli for the same purpose some 300 or more years ago.

In figure 15.3 is represented a bulb syringe, S, valved at A, and having a short tube, B, which dips into a basin of water. Leading from the opposite pole of the bulb is a longer tube, C. When the bulb is alternately compressed and released fluid is drawn from the basin and discharged from

the mouth of the tube. If the walls of the latter are composed of some rigid material (fig 15.3, 1), it will be found that when the pump is worked the fluid issues from the tube in spurts or jets synchronous with each stroke, but no flow occurs between the strokes. An increase in the frequency or force of the strokes does not alter the intermittent character of the flow nor does lengthening the tubing. If the peripheral resistance of the vascular system be imitated by attaching a nipple of small bore to the mouth of the delivery tube so as to increase the resistance to the outflow of fluid, the issuing stream is finer and its velocity is increased but it still remains intermittent (fig 15.3, 2). Let the elasticity of the arterial wall now be imitated by replacing the rigid tube by one of rubber, yet let the mouth of the tube be left free and not constricted in any way (fig 15.3, 3). The intermittent character of the stream is unaffected. However, if the small-bored nipple representing the peripheral resistance be fixed into the mouth of the elastic tubing the stream will be found to have lost its pulsatile character and to have become continuous (fig 15.3, 4). Two factors are therefore necessary to pro-

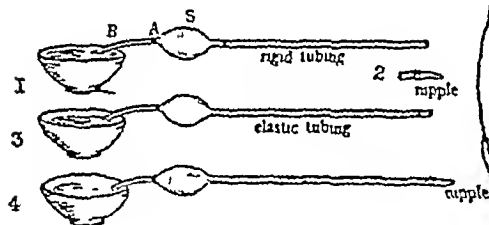


FIG 15.3 Description in text.

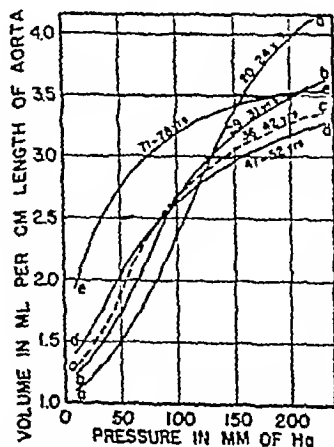


FIG 15.4 Curves of the absolute volume of liquid per unit length of human aortas at different ages at pressures from 10 to 230 mm Hg. *a*, age 20 to 24 years, *b*, 29 to 31 years, *c*, 36 to 42 years, *d*, 47 to 52 years, *e*, 71 to 78 years. (After Hallock and Benson)

duce this result, (a) *resistance to the outflow* and (b) *elastic tubing*. The reasons for this are clear. If the fluid has free egress from the tube most of that which enters it from the pump is discharged from the open end before the next beat occurs, the pressure, in consequence, does not rise to a sufficient height to distend the rubber wall, i.e., elasticity is not called into play, over-filling of the tube does not occur, and in consequence the latter acts simply as though it were composed of rigid material.

The foregoing facts apply directly to the arterial system. The elasticity of the vascular walls and the peripheral resistance are both essential for the maintenance of the diastolic pressure. As the contents of the ventricle are thrown into the already over-filled system during systole the added pressure which is then exerted upon the vascular walls causes their further distension. After the completion of systole the elastic walls rebound and, pressing upon the blood within their embrace, force it onwards through the peripheral vessels. In other words, the arterial lumen returns to its previous diameter and the energy that had been stored up during the stretching of the elastic tissue is in this way gradually expended during diastole.

The elastic recoil of the arterial wall thus acts in a sense as a subsidiary pump to drive the blood onwards in a continuous stream between the heart beats. Otherwise the pressure would fall to zero after each systole.

It is clear then that any reduction in the elasticity of the arteries, other factors remaining unchanged, will tend toward a lowering of the diastolic pressure. If the aorta and its larger branches are stiffened (as a result of sclerotic changes), they cannot expand to the same degree as can healthy, resilient vessels, and therefore do not so readily accommodate the blood (60-100 cc.) ejected from the heart during systole. Such a state will lead to a rise in the systolic pressure. Normally, the cross sectional area of the aorta increases considerably with age, that is, its capacity, except at higher pressures, is greater than in early life, but its distensibility decreases (fig 15.4) steadily after about the 25th year.

THE INFLUENCE WHICH VARIATIONS IN SOME OF THE FOREGOING FACTORS EXERT UPON THE DIFFERENT PHASES OF THE ARTERIAL PRESSURE

1. *Change in heart rate unaccompanied by an alteration in any of the other factors, e.g., output of heart per minute (p. 263), peripheral resistance, etc., will*

cause a change in the diastolic pressure but relatively little change in the systolic. During cardiac acceleration, for example, the diastolic period is shortened and less time is therefore allowed for the energy stored in the elastic walls during systole to become converted into energy of flow during diastole. In other words, the fall in pressure during diastole is halted at a higher level by the earlier arrival of the next beat. A decrease in heart rate will have the opposite effect, with the longer diastole the slope of pressure is enabled to reach a lower level. Since the quantity of blood entering the arteries per minute remains constant the quantity entering at each beat must vary inversely with the change in rate, which accounts for the relatively small change in the systolic pressure.

(2) *Alterations in the quantity of blood discharged per minute by the ventricle.* If little change should occur in the heart rate and other factors remain unaltered, increase in the output per beat of the heart causes a rise chiefly of the systolic pressure. The diastolic pressure is raised less noticeably, consequently the pulse pressure is increased. The explanation for the less pronounced rise in the diastolic pressure is that, as a result of the high pressure at the end of the ejection period, the pressure gradient throughout diastole is steeper and more energy is expended in giving velocity to the blood, of the blood pumped into the arterial system during systole a larger proportion than ordinarily will therefore have passed through the arterioles by the end of diastole.

(3) *Changes in the peripheral resistance* while other factors remain constant. Though these affect both systolic and diastolic pressures they show their influence predominantly upon the latter phase. The diastolic period is considerably longer even in a rapidly beating heart than the ejection period of systole (p 208) and, as we have seen, the peripheral resistance is an important factor in the maintenance of the diastolic pressure. It follows therefore that any increase or decrease in the outflow from the arterial system will affect this pressure to a greater degree than the systolic. The mean pressure and pulse pressure vary accordingly. Aortic regurgitation produces an effect upon diastolic pressure similar to that caused by a reduction in the peripheral resistance but greater in degree. The mechanical principles involved are similar, an increased quantity of blood passes from the arterial system during diastole as a result of leakage through the incompetent aortic valves. The peripheral vessels are also usually dilated, which, combined with the high pulse pressure, may cause the appearance of a pulse in the capillaries especially if their emptying be aided by holding the arm above the heart level. Slight pressure upon a superficial capillary area such as at the base of the finger nail may then show alternate blanching and flushing synchronous with the heart beat (see also p 215). In aortic regurgitation the pulse pressure attains a magnitude seen in no other

condition (80 or 110 mm Hg) for not only is the diastolic pressure much reduced but the systolic pressure is raised as well, owing to the greater volume of blood ejected at each heart beat—that which has regurgitated through the aortic valves plus that received from the auricle. For these reasons, the carotids throb visibly. The pulse is of the collapsing type (water hammer or Corrigan's pulse) (p 215). An arterio-venous aneurysm produces somewhat similar effects (p 266) upon the arterial system.

(4) *A rise or a fall in blood viscosity*, other factors remaining unchanged, tend to affect the diastolic pressure in a manner similar to changes in the calibers of the peripheral vessels.

(5) *Increase in blood volume* will raise both pressures, as a result of the overfilling of the arterial system and greater stretching of the elastic walls.

(6) *Reduction in elasticity of the arterial walls.* Obviously a condition such as arteriosclerosis, which renders arteries less resilient and more like rigid tubes will, tend toward a lowering of diastolic pressure (p 149). Yet as a matter of fact, in arteriosclerosis the diastolic pressure may be raised rather than lowered, since there is frequently an associated narrowing of the peripheral vessels which more than offsets the hardening of the walls of the larger arteries. When, however, the sclerosis is confined to the larger vessels and their branches while the peripheral vessels are free from proliferative changes which narrow their lumina the diastolic pressure is lowered. Diminished distensibility of the walls of the aorta and the larger arteries tends to increase the systolic pressure.

In the foregoing paragraphs variations in the several blood pressure factors and their effects have each been considered as being the only variable in a particular instance. The object of this was to disclose the value of each and the manner in which it acted. Yet it must be remembered that such a description is more or less artificial and that in health and even to a large extent under pathological conditions the various factors interact with one another—there is a give and take among them. When a change in the value of one factor occurs, readjustments of others take place to regulate the blood pressure and keep it within the normal limits. For instance, dilatation of the vessels in one area may be accompanied by vasoconstriction in another (p 288). Reduction in blood volume as by hemorrhage, is followed by constriction of the peripheral vessels (p 28), while increased blood volume or a rise in viscosity will likely be followed by the opposite effect upon the vessels.

THE BLOOD PRESSURE IN THE HUMAN SUBJECT

THE VENOUS BLOOD FLOW

METHODS OF MEASUREMENT

Though the blood pressure can be determined in man by the insertion of a hollow, wide-bore needle into an artery and connecting it with a suitable manometer, such a method is usually reserved for experimental work. Convenience and safety demand that any method for general clinical use must be indirect. The principle employed consists in balancing air pressure against the pressure of the blood in the brachial artery and then estimating the former by means of a mercury or an aneroid manometer

The instrument used for this purpose includes a flat rubber bag covered by an undistensible envelope of cotton fabric. The cavity of the bag is connected by a length of rubber tubing with the manometer and by another tube with a hand bulb or small pump. By this means the bag can be inflated to any desired pressure. A small valve between the bulb or pump and the bag permits the escape of air, and the reduction of the pressure as required. The uninflated rubber bag (usually referred to as the cuff or armlet) which should be at least 12 cm. wide is wrapped snugly around the upper arm just above the elbow. The bag is then inflated until the air pressure within it overcomes the arterial pressure and obliterates the arterial lumen. The pressure is increased a little beyond this point and is then cautiously reduced again,¹ by the release of the valve, until the arterial pressure just overcomes the air pressure and blood escapes beyond the cuff into the peripheral section of the artery. At this instant the pressure in the bag is read from the manometer. Since the air pressure practically balances the systolic arterial pressure the manometer reading must indicate the value of the latter. It is essential to the method that the manometer reading be taken at the instant when the blood escapes beneath the cuff.

One of three methods, the *palpatory*, *oscillatory*, or *auscultatory*, may be employed to determine the latter. In all three procedures the value for the lat-

¹Should deflation be carried out too rapidly the mercurial type of clinical manometer shows a considerable lag, due to the time required for the air above the mercury column to reach atmospheric pressure. The rate of deflation should not exceed from 2 to 3 mm. Hg per second.

eral pressure in the brachial artery is obtained, whereas direct methods, which entail the introduction of a straight, unbranched cannula, or a large bore needle, into an artery, measure the end pressure. The latter is greater because it includes not only the lateral pressure on the elastic arterial wall, but also that derived from the conversion of the kinetic energy of the moving blood column as it meets the obstruction, as well as an increment of pressure caused by the reflected wave from the obstruction.

The palpatory method. In this method the examiner takes the moment that the pulse is felt at the wrist as the index of the systolic pressure. This method is now rarely used since it lacks accuracy. It assumes that the first escape of blood beneath the cuff will cause pulsation in the peripheral artery, but there is no evidence that the amount of blood which escapes beneath the cuff when the artery first opens is sufficient to produce a pulse wave detectable by the finger. Definite pulsation may not occur until the cuff pressure has been lowered 5 to 10 mm. below the point when the artery first becomes pervious. This method therefore gives readings that are too low. Another disadvantage of the palpatory method is that the diastolic pressure cannot be measured satisfactorily.

The oscillatory method. In this method a tambour or capsule covered with a very delicate membrane, or a second bag connected with the cuff, is used to pick up and magnify the pulsations transmitted from the artery to the upper edge of the latter. The pulsations are made to appear as oscillations of the indicator needle on the clock face dial of an aneroid manometer. Pressures are marked by figures on the dial, as the cuff is inflated or deflated, the needle moves to indicate the applied pressure at the moment. At pressures exceeding systolic the oscillations are minimal, but as the pressure is gradually lowered and the pulsations pass beneath the cuff, a sudden increase in their amplitude and duration occurs. This is the criterion of systolic pressure. The oscillations show little change in magnitude as the cuff is deflated further, until the pressure has fallen to the diastolic level at which they suddenly become smaller. At this instant the figure on the dial to which the needle points is noted.

THE AUSCULTATORY METHOD

This procedure is the one generally employed clinically. It was introduced in 1905 by the Russian physician Korotkow. Certain sounds heard during auscultation of the brachial artery below the cuff are taken as the criteria for the systolic and diastolic pressures. Under ordinary circumstances if a stethoscope is placed upon the brachial or any other artery, no sound can be heard, the flow of blood along the arterial channels being inaudible. If however the artery is compressed by the manometer cuff so as to completely arrest the flow of blood for a moment, a sharp light tapping sound in rhythm with the heart beat will be heard when the pressure in the cuff is again released and falls just sufficiently to permit the arterial lumen to open and allow a jet of blood to pass beyond. As the pressure in the cuff is progressively lowered the sound undergoes a series of changes in quality and intensity.

Four phases of the sound, each having its distinctive character, may be heard in succession in the normal individual, as the pressure is gradually reduced from about 120 to 80 mm. of mercury or less. These are given below with the average pressures at which they are normally heard.

Sounds of Korotkow

- Phase I Sudden appearance of a clear, but often faint, tapping sound growing louder during the succeeding 10 mm. Hg fall in pressure.
- Phase II The sound takes on a murmurous quality during the next 15 mm. fall in pressure.
- Phase III Sound changes little in quality but becomes clearer and louder during the next 15 mm. fall in pressure.
- Phase IV Muffled quality lasting throughout the next 5 to 6 mm. Hg fall. After this all sound disappears.

The beginning of the first sound is taken as the *index of systolic pressure*. As it is quite faint at its commencement it may not be caught at this time by the ear of the inexperienced, or if the observer's hearing is distracted by other sounds. The sound then will not be picked up until the pressure has dropped below the level at which it could be heard in quiet surroundings and the reading of the systolic pressure will be too low.

The fourth sound just before its complete disappearance is taken as the *index of the diastolic pressure*. This sound coincides with the moment

that the blood escapes beneath the armlet in a continuous stream rather than intermittently.

The measurement of blood pressure in man is attended by certain inaccuracies dependent upon the resistance of the tissues of the part and, under certain conditions, of the arterial wall itself to the compressing force. These fallacies are to a large extent mitigated by the use of a broad armlet (12 cm. in width), as first introduced by von Recklinghausen, which distributes the applied pressure over a wide area. In a person with normal arterial walls the pressure as measured is probably very close to the true systolic pressure. Variations in the resistance of the arterial wall in different individuals as a result of sclerotic changes or simple hypertonus of the muscular coat may lead to error and give readings that are too high. Repeated compression and decompression just before the actual determination is made will, as a rule, soften the artery or remove any spasm of its walls sufficiently to eliminate this source of inaccuracy. Even when definite arteriosclerosis exists, approximately correct estimations of the blood pressure are obtained when this precaution is taken. Though lower readings as a rule are not obtained in a normal individual by repeated readings, in others with apparently normal arteries the reading obtained after the third or fourth trial may as a result of the reduction in tone of the vessel wall be lower than the initial observation by several millimeters.

RELATIVE VALUES OF SYSTOLIC AND DIASTOLIC MEASUREMENTS

The systolic pressure is subject to wider variations under ordinary conditions of health than the diastolic, it also varies more with local arterial changes, for these reasons less reliance can be placed upon it than upon the diastolic. A knowledge of the diastolic level is also of greater value for other reasons. (1) It represents the constant load which the vascular walls are carrying, not only in the larger arteries but throughout the arterial system. (2) It reflects more accurately the state of the peripheral vessels, for as already pointed out (p. 151) the systolic pressure responds less to variations in the peripheral resistance. (3) It determines the filling of the coronary system (p. 326).

THE RELATION OF THE PULSE PRESSURE TO THE CARDIAC OUTPUT

The pulse pressure, other factors remaining the same, varies with the amount of blood ejected into the aorta per beat, and we know that the work of the heart depends chiefly upon the amount of blood ejected per minute against the mean pressure in the aorta. It fol

lows therefore that provided the peripheral resistance, blood viscosity and other factors concerned in the maintenance of blood pressure remained constant the product of pulse pressure and heart rate ($P \times H \times R$) might be used as an index of the energy of the ventricular contraction. Erlanger and Hooker found that the product tended to remain constant under ordinary physiological conditions, which would not be expected to alter the cardiac output, an increase or decrease in heart rate occurring to compensate for changes in pulse pressure. The index in an individual having the usual pulse pressure of 40 and a pulse rate of 72, would be ($40 \times 72 =$) 2880. Conditions causing a pronounced rise or fall in the cardiac output, on the other hand, are accompanied by a corresponding change in the product. The index has therefore been used as a rough clinical gauge of the minute volume of the heart, and so as a guide to the manner in which the heart is performing its work in diseased states. The index may be of value in following the course of an individual case if the observations are made at comparatively short intervals and the conditions to which the patient is exposed remain practically the same from one observation to the next. No absolute or quantitative value can, however, be attached to it, nor can the results obtained in different individuals be compared with one another.

THE NORMAL ARTERIAL BLOOD PRESSURE

The average systolic pressure of young male adults at mental and physical rest in the sitting position (i.e., as usually measured in a routine medical examination) is usually given as 120 mm Hg, the diastolic as 80, the mean pressure as 100 and the pulse pressure as 40. The pulse pressure is therefore one half of the diastolic and one-third of

the systolic, and the normal ratio of systolic, diastolic and pulse pressures is 3-2-1. This ratio holds only at these levels, being altered if the pressures are much above or below the foregoing figures. Robinson and Brucer conclude from a study of nearly 11,000 persons that the range of normal blood pressures is from 90 to 120 mm Hg systolic and 60 to 80 mm Hg diastolic. They contend that blood pressures above the upper levels just mentioned cannot be considered as physiological at any age.

Slight diurnal variations in blood pressure of from 5 to 10 mm. Hg systolic occur, the peak being in the afternoon and the lowest level in the early hours of the morning.

VARIATIONS IN THE BLOOD PRESSURE UNDER PHYSIOLOGICAL CONDITIONS

Age, sex, and build

Age exerts a definite influence upon the blood pressure levels. At birth the systolic pressure measures from 20 to 60 mm with an average of 40 mm. It rises rapidly, however, and has an average value of about 70 mm. at the end of a fortnight and 80 mm. at the end of a month. A slow steady rise takes place from this time until about the twelfth year when it averages 105 mm. With the onset of puberty a more sudden rise occurs, which in boys reaches 120 mm. at about the age of 17. In girls, there is an increase in systolic pressure to the fifteenth year, then a decline to the eighteenth, it remains fairly steady from then on or shows a gradual rise. It has usually been considered that a steady though not great rise in blood pressure from adolescence to old age is the rule in health, the averages for the age of 60 being given as about 140 systolic pressure and 87 diastolic. In men, it is usually stated that a rise of about 0.5 mm. in the systolic and 0.2 mm. in the diastolic and pulse pressure occurs for each year of age after 20 (see table 16) but, from the observations of Robinson and Brucer, mentioned above, it is debatable whether this increase with age can be considered to be a physiological phenomenon. In women up to the time of the menopause the systolic pressure is from 4 to 5 mm. lower than for men of the same age. At the menopause, however, there is a somewhat abrupt rise and the pressure remains a little above the male average from then onwards.

TABLE 16

The average variations of blood pressure (after Hunter's compilation of observations on a quarter million healthy Americans)

(After Gager)

AGE	SYSTOLIC PRESSURE	DIASTOLIC PRESSURE	PULSE PRESSURE
10	103	70	33
15	113	75	38
20	120	80	40
25	122	81	41
30	123	82	41
35	124	83	41
40	126	84	42
45	128	85	43
50	130	86	44
55	132	87	45
60	135	89	46

Symonds and others have found a correlation between the systolic pressure and obesity. Comparing groups of overweight and normal individuals the former were found to have a pressure on the average 7.5 mm higher than the latter. In markedly obese but otherwise healthy persons the difference was even more pronounced. The incidence of abnormally high blood pressure (hypertension) is also definitely greater in persons of overweight.

Robinson and Brucer claim that *body build* is correlated with the blood pressure level. They found in an examination of a large number of persons that in any weight group broad-chested persons on the average had a higher blood pressure (both systolic and diastolic) than had those of slender build.

The effects of digestion, emotion, exercise and posture

DIGESTION The systolic pressure is influenced to a small but definite extent by meals. A rise of from 6 to 8 mm is the usual effect, and this lasts for an hour or so. There is little change in the diastolic pressure, if anything it is reduced, presumably a result of vasodilation in the digestive organs and skin.

EMOTIONAL INFLUENCES, excitement, fear, worry, etc. markedly affect the arterial blood pressure, especially the systolic. The effects are brought about through increased cardiac action and changes in the state of the vessels through impulses playing upon the cardiac and vasomotor centers in the medulla. The liberation of adrenaline into the blood stream may also be a factor. Quiet restful *sleep*, according to MacWilliam, is accompanied by a fall of from 15 to 30 mm in the systolic pressure. The fall is most marked during the first hours, rising gradually again after this until the time of waking. MacWilliam observed that if the sleep was disturbed and accompanied by imaginary motor activities there might be no depression of the pressure, but rather an elevation, in some instances to as high as 200 mm systolic and 105 mm diastolic.

EXERCISE Of all physiological conditions this, if of a strenuous nature, has the most powerful effect upon the arterial blood pressure. During the muscular effort or even immediately before, i.e., at the instant that the exertion is contemplated, the pressure commences to rise and reaches a height of 180 or 200 mm. The diastolic pressure shows a less pronounced rise (100 to 110) so that

the pulse pressure is increased. In light exercise the diastolic pressure may remain at the normal level while the systolic rises several millimeters. Immediately after the exercise the pressure drops momentarily to normal or even slightly below. It then mounts rapidly to its previous high level, from which it gradually declines again, and in a healthy person reaches the normal within from 1 to 4½ minutes. The evanescent drop in pressure is explained by Cotton, Lewis and Rapport as being due to the sudden relaxation of the abdominal muscles. The blood is drained into the venous reservoirs. These when deprived of their support (abdominal muscles) have their capacity increased and the blood flow into the right heart is temporarily curtailed. It is not until an appreciable time has elapsed to enable the increased venous capacity to become filled again by blood pouring in from the recently active muscles that an adequate flow into the right heart is restored (p. 168). These responses are much more pronounced and the final decline of pressure to normal levels is postponed if the exercise is carried out in a rarefied atmosphere, or if a condition of "irritable heart" exists.

POSTURE The diastolic pressure is somewhat higher in the standing than in the sitting position and lowest in recumbency. This change is found to occur whether the postural change is brought about actively or passively and is evidently an over-compensation for the gravity effect (p. 169). The systolic pressure usually rises but to a less extent than the diastolic, so the pulse pressure is reduced. It is difficult to obtain a blood pressure reading in less than 30 seconds, but those which have been taken at as short an interval as 10 seconds after the erect position has been assumed show that the initial effect is a fall of from 6 to 22 mm Hg in the systolic pressure. This is sufficient to stimulate the carotid sinus and aortic mechanisms and cause increased vascular tone with a consequent compensatory rise in pressure. Compensation is usually complete within 30 seconds. Reverting from the standing to the sitting or recumbent position has the reverse effect, fall in diastolic pressure and rise in pulse pressure. In persons with an abnormally and habitually low blood pressure, the systolic pressure may actually rise in the lying-down position and fall when the subject stands. The diastolic, on the other hand, is always lowered in recumbency and raised in the erect posture.

PATHOLOGICAL VARIATIONS IN THE ARTERIAL BLOOD PRESSURE

The blood pressure may be persistently above or below the normal range. These departures from the normal are termed *hypertension* and *hypotension*, respectively. It is difficult to make a sharp separation of the normal from the abnormal but an elevation above the average normal for a particular age, of 15 mm in the systolic and 8 mm in the diastolic may be considered to be definitely abnormal. A reduction below 110 mm in an adult male (or 100 mm in adult females) of any age is usually termed hypotension.

ARTERIAL HYPERTENSION

A high blood pressure accompanies such conditions as *increased intracranial pressure*, *hyperthyroidism* and *adrenal tumor or hyperplasia* (ch 59). In the first two of these, the hypertension is more or less incidental or of subordinate importance to the primary disease. In the first the heightened pressure is apparently due to a generalized vasoconstriction resulting from the reduced oxygen supply to the vasomotor center. Cushing pointed out that the increased pressure within the cranial cavity caused compression of and slowing of the blood flow through the vessels supplying the medulla. In animals, sudden occlusion of the common carotids, after the vertebrals and other branches of the subclavian have been ligated, causes a pronounced rise in blood pressure. Even though the carotid sinus has been excised, chronic hypertension may be produced in dogs by cerebral ischemia produced in a similar way. Also, a pronounced rise in blood pressure may be induced by asphyxia (ch 27), the increased hydrogen ion concentration of the blood acting as the stimulus to the vasomotor center. Permanent hypertension has been produced by the injection of an inert substance such as kaolin into the cisterna magna of rabbits (Dixon and Heller) and thus interfering with the blood supply to the medullary centers, and cerebral anemia produced by ligation of the carotids and vertebrals causes hypertension in the dog. In hyperthyroidism the hypertension is associated with an increased cardiac output.

Temporary rises in blood pressure also occur in attacks of *angina pectoris*, in *lead colic* and in the *crises of tabes*. The hypertension associated with adrenal tumor or hyperplasia is usually of a paroxysmal character and is due to liberation of ex-

cessive amounts of the medullary secretion. (ch 59)

The types of hypertension which are the main consideration of this section are, (1) *hypertension secondary to renal disease*, (2) *experimental hypertension*, and (3) *essential or primary hypertension*.

HYPERTENSION SECONDARY TO RENAL DISEASE

Sooner or later in chronic glomerulo-nephritis the arterial blood pressure rises, the left ventricle hypertrophies and the arterial tree shows degenerative changes (atherosclerosis).

PATHOGENESIS It is a well-attested fact that in this form of the disease as well as in primary or essential hypertension increased peripheral resistance is the immediate causative factor. The increased resistance is mainly in the splanchnic area (Abramson). None of the other factors which have been enumerated in the last chapter as sustaining the normal blood pressure exerts an excessive influence in these hypertensive states. The hypertrophy of the heart, for example, is purely secondary—a physiological compensation rendered necessary, as first suggested by Richard Bright (1827), by the greater resistance offered to the flow of blood through the peripheral vessels.

Divergent views have been held as to the manner in which the increased peripheral resistance is brought about. Of the theories which have been proposed two require some mention. A brief account will serve as a background to a description of modern researches upon the subject.

(a) *Reduction in size of the vascular bed of the kidney* as a result of structural narrowing of the renal arterioles and the destruction of large numbers of glomerular capillaries. Such changes, it was presumed, would narrow the outlet from the arterial to the venous system. There is no evidence that a localized increase in peripheral resistance of this nature occurs and is responsible for hypertension, for in most animals (rat excepted) removal of one kidney and part of the other has little or no effect upon the blood pressure. Anderson reduced the renal tissue in rabbits to the point where severe renal insufficiency resulted without causing any rise in blood pressure. From the great adaptability of the circulation, and the enormous potential capacity of the peripheral vessels, this is what one should be led to expect, the removal of a vascular area from one part of the body being readily compensated for by the opening of vascular channels elsewhere. Reducing the vascular bed of the kidney by the ligation of vessels does, however, cause hypertension if the renal tissue supplied by the occluded vessels is left *in situ*.

(1938) obtained a pressor substance from normal issue of rabbits. They postulated that defecation of this material in kidney disease and its retention in the circulation was responsible for a generalized vasoconstriction. The fact that renal insufficiency caused by the removal of renal tissue is not a feature of hypertension is in harmony with such a theory. For the production of a pressor substance by the kidney would also be reduced by the operation. On the other hand, Geer and Dragstedt found that in which the urine had been diverted into the small intestine or directly into the blood showed no elevation of blood pressure, though they were observed for periods ranging from 11 days to 5 weeks. As will be seen presently, the conclusions of Tjallingii and Bergman came remarkably close to modifying the thought concerning the production of hypertension in renal disease.

PRODUCTION OF HYPERTENSION BY EXPERIMENTAL CONSTRICTION OF THE RENAL ARTERY

Though others had reported that compression of the renal arteries in animals caused a rise in blood pressure, Goldblatt and his associates were the first to show that this procedure causes consistent, pronounced and permanent hypertension. Constriction of the renal artery is effected by the use of a specially devised adjustable silver clip. Ischemia of one kidney, produced in this manner while the other remaining intact, causes a moderate elevation of the blood pressure which commences within four days after the operation, but persists for a short time, returning to the normal level within a month or so. If, on the other hand, both renal arteries are constricted or if one alone is constricted and the opposite kidney removed, permanent hypertension results. The severity of the hypertension varies with the degree to which the renal blood flow has been curtailed. Unless the constriction of the renal artery is extreme and, in consequence, the ischemia of the kidney very severe, the elevation of the blood pressure is not accompanied by any detectable impairment of renal function and the kidney shows little histological change. But with severe constriction of the renal artery, renal insufficiency develops and the animal may die in uremia.² The hypertension and

degenerative changes in the systemic arterioles, the arteriolar walls in many instances showing hyaline degeneration and necrosis. Such a state of the vascular system, taken together with renal insufficiency and the elevated blood pressure, is closely comparable to severe clinical hypertension. But Goldblatt was unable to produce arteriolar degeneration in the kidney itself for the reasons that, (a) a severe degree of hypertension could not be produced as long as one non-ischemic kidney remained in place, and (b) the vessels of the ischemic kidney are protected from the destructive effect of the hypertension by the compressing clamp, i.e., they lie within a region of low pressure. Wilson and Byrom have succeeded, however, in producing severe hypertension (up to 260 mm Hg) in rats by constricting one renal artery and leaving the opposite kidney *in situ*. Severe vascular damage, consisting of hyaline or fibrinoid degeneration and necrosis of the walls of the arterioles of the intact kidney (especially of the afferent glomerular vessels) as well as of the systemic vessels, was observed.

THE MECHANISM INVOLVED IN THE PRODUCTION OF HYPERTENSION BY RENAL ISCHEMIA. RENIN. It has been definitely established that the hypertension following constriction of the renal artery is the result of a substance produced by the ischemic kidney. The hypertensive effect is not dependent upon a reflex through afferent endings in the kidney, the vasomotor center and vasoconstrictor fibers to the systemic vessels for it occurs after

removal of the kidney, is not followed by elevation of the blood pressure, nor does constriction of the renal vein result in more than a very transient rise in pressure. Hypertension also follows constriction of the aorta above the origins of the renal arteries. Page found that hypertension develops within from four to six weeks after a kidney has been wrapped *loosely* in cellophane or silk. These materials set up a chronic inflammatory reaction (perinephritis) which leads to the formation of a firm fibrocartilagenous capsule enclosing the kidney. The hypertension is the result, apparently, of the renal ischemia induced by the compression of the vessels within the kidney by the newly formed tissue. Contemporaneously and independently, Greenwood, Nassim and Taylor observed that if one kidney was surrounded by a closely fitting cast of gauze stiffened by impregnation with collodion, hypertension occurred *within a few hours after the removal of the opposite kidney*, progressed in severity and persisted indefinitely. No rise occurred as long as the other kidney remained undisturbed. The mechanism bringing about the hypertension by this method, therefore, appears to be different from that following the loose envelopment of the kidney in cellophane or silk, for it occurs before a constricting perinephritis could have developed.

denervation of the kidney, or section of the splanchnic nerves or of the anterior spinal nerve roots. Even complete excision of the sympathetic chains or destruction of the spinal cord does not prevent the development of the hypertension. Furthermore, hypertension results if one kidney which has been transplanted into the inguinal region or into the neck (renal artery to carotid, renal vein to jugular) is made ischemic and the other kidney then excised. The transplanted kidney is, of course, completely isolated from nervous control. The inability of such denervation operations to prevent or modify the hypertensive effect proves conclusively that it is mediated through a pressor substance circulating in the blood stream and, further, that the latter acts, not through the vasomotor center, but either directly upon the peripheral vessels or through the intermediary of one or other of the endocrine organs (see below).

The humoral substance formed in the ischemic kidney which is primarily responsible for the hypertensive effect is called *renin*, the name already mentioned as having been originally used by Tigerstedt and Bergman. Renin, as we shall see presently, produces a vasoconstrictor material (see hypertension, below) but is itself free from any direct pressor action. It is a proteolytic enzyme, acting best at a pH between 7.5 and 8.5 and at a temperature of about 37.5°C. Its substrate is an alpha globulin of the plasma. It is non-dialyzable and thermolabile being destroyed above 56°C, and is probably a protein. The renin of pig, dog, monkey, ape or man produces its characteristic effect

when injected into a lower mammal, and that of the monkey, ape or man when injected into any one of these primates. But it shows a special type of specificity, for the renin of a lower mammal is inert when administered to monkey, ape or man. The renin of the bird is active only in birds.

It is probable, as Goldblatt believes, that renin is present in the renal tissue in the form of a precursor (*crypto renin*), and is activated only after it has been released into the circulation.

Difficulty has been experienced in demonstrating in a direct way the presence of a pressor substance in the circulation of the hypertensive animal. The blood pressure of a normal dog is not elevated by transfusing it with blood drawn from one with experimental hypertension. Extracts of ischemic kidneys have yielded uncertain results, for renin is present in the normal kidney though in somewhat smaller amounts, it is claimed, than in the ischemic organ. Transfusion and crossed circulation experiments have yielded conflicting results. Katz and his colleagues transfused non hypertensive nephrectomized dogs with the blood of hypertensive animals over a period of 18 hours but observed no rise in the blood pressure of the recipient animals. On the other hand, Solandt, Nassim and Cowan have reported that when by means of a specially designed pump large volumes of blood were transfused from a hypertensive to a non hypertensive nephrectomized animal, a temporary rise in blood pressure resulted.

Nevertheless, convincing evidence of the production of a hypertensive substance in the ischemic kidney was furnished by the experiments of Housay, Tasciolo and Taquim. They found that the blood pressure of a non hypertensive, nephrectomized animal was raised when the ischemic kidney of a hypertensive animal was transplanted into its neck, whereas the transplantation of the kidney of a normal animal gave a negative result. They also showed that the plasma of blood collected from the vein of an ischemic kidney, as compared with that from a normal one, caused a pronounced vasoconstrictor effect when perfused through the vessels of the toad.

Renin has been reported to be present in the blood of patients with acute hypertension, but not in cases of long standing. It has also been found in the blood of persons poisoned by barbiturates, as well as in the blood of animals with renal hypertension (fig 16.1). A rise in the concentration of renin in the blood has been observed in animals

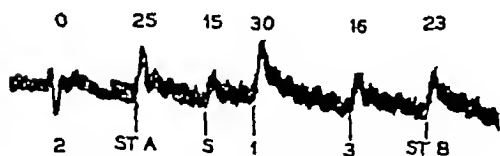


FIG. 16.1 Demonstration of renin in the blood of a patient with acute glomerulonephritis and hypertension. Blood pressure curves are those of a cat under dial anesthesia. The figures above the curves indicate the rise in blood pressure in mm Hg. 2—Response to injection of unincubated control equivalent to 10 cc serum. This response, it will be observed, is about the same as that given by the saline control (1). St. A—Response to 0.75 cat unit of standard hypertensin. S—Response to 10 cc physiological saline. 1—Response to incubated sample equivalent to 5 cc serum. When calculated for renin, this is equivalent to 2.4 cat units of renin per 10 cc. serum. 3—Response to sample incubated with hypertensinase. Note that this gives approximately the same rise as the saline control. St. B—Response to 1.2 cat units of standard hypertensin (after Dexter and Haines).

suffering from surgical shock, following hemorrhage and in other states associated with a profound fall in blood pressure. Such observations suggest that the liberation of renin by the normal kidney is a physiological factor in the regulation of the blood pressure, aiding in its restoration to higher levels when it becomes greatly lowered.

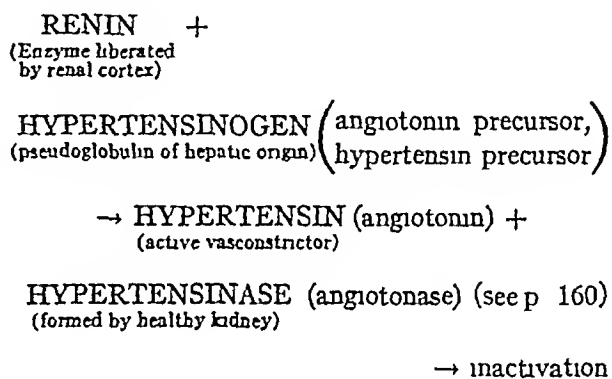
THE ACTIVE VASOCONSTRICTOR AGENT Helmer and Page found that the more purified preparations of renin were inactive. When perfused with Ringer's solution through the vessels of an isolated organ such as the rabbit's ear or dog's tail, such preparations showed little or no vasoconstrictor action. The addition of plasma, blood or serum globulin to the purified renin preparation restored its vasoconstrictor property. Blood or plasma, therefore, contains a material which Helmer and Page called *angiotonin precursor*, but is now more commonly known as *hypertensinogen*, the name given to it by Braun-Menendez and his associates. This substrate of renin is believed to be a pseudoglobulin (alpha₂ globulin of plasma). The active vasoconstrictor material produced by the action of renin is called *hypertensin* (Braun-Menendez) or *angiotonin* (Page). The hypertensive action following the injection of renin becomes less pronounced with repeated injections. This tolerance to repeated injections of renin has been termed *tachyphylaxis*. The phenomenon is attributed by Page and his associates in part to the exhaustion of hypertensinogen. The serum globulin is produced, apparently, in the liver since it disappears from the blood after hepatectomy. It is said to be increased in experimental hypertension and in clinical essential hypertension.

Chemical and physiological properties of hypertensin

Hypertensin is believed to be a polypeptide of low molecular weight, namely, 2700. According to Edman's analysis of a highly purified preparation, it contains 28 per cent of histidine and small quantities of lysine, glycine, alanine, serine, proline, valine, leucine, tyrosine, and aspartic and glutamic acids. Hypertensin is dialyzable and thermostable, it is not species specific, preparations from animal kidneys producing a pressor effect in man and in all species investigated. One milligram of Edman's purified preparation has a hypertensive action equal to 39 mg of tyramine phosphate, it is therefore comparable in potency, so far as its blood pressure raising effect is concerned, with commercial adrenaline. But it constricts the coronary

arteries and reduces cardiac output and efficiency. It reduces splenic and renal volumes and increases venous pressure, the cutaneous blood flow is unaffected. Immediately after its administration the blood pressure rises steeply, but the rise is short-lived, its action thus differing from that of renin, which is gradually developed (characteristic dependent upon the delay required for the production of the pressor principle itself) and prolonged. It acts directly upon the vascular musculature and also stimulates nearly all other smooth muscle. It reduces renal blood flow apparently through constriction of the efferent glomerular arteriole. Glomerular pressure is thus raised and the proportion of fluid filtered from the plasma increased.³

The renin-hypertensin mechanism may now be summarized in tabular form, thus —



THE SOURCE OF RENIN AND THE ADEQUATE STIMULUS FOR ITS LIBERATION Renin is confined to the cortex of the kidney. But whether it is formed in the cells of the tubules, or in the juxtaglomerular apparatus as claimed by Goormatigh, is not definitely known. Hypertrophy of the cells of these structures has been described in hypertensive animals, and others report that, in dogs, they contain granules which exhibit histochemical properties similar to those of renin, and that there was a very good correlation between the number of these granules and the renin content of the kidney (Marshall and Wakerlin). On the other hand, experiments with a tubular poison (tartrate) point to the cells of the tubules as a more likely site of renin production. Kidneys in which the cells of the tubules have been destroyed by tartrate, do not contain renin, whereas kidneys of which thin wafers of the renal cortex were shaved off (Braun-

³ A vasoconstrictor substance (*pepsitensin*), closely resembling hypertensin in its effects, has been obtained by Croxatto and associates by the action of pepsin upon plasma globulin.

Menendez), to remove the glomeruli and the juxta-glomerular apparatus, still contained the enzyme.

The stimulus for the production or liberation of renin by the kidney is not known. It is denied by Page that ischemia of the kidney is itself responsible, for, though the immediate effect of constriction of the renal artery is a fall in pressure in the vessels of the kidney, and a reduction in blood flow, later on the renal blood flow may return to normal. Nor did anoxia of the renal tissue seem to be a necessary condition for the development of renal hypertension, because no increase over the normal was found in the arterio-venous oxygen difference of the ischemic kidney (Levy and associates). Moreover, anoxia induced by cyanide does not cause renin to appear in the circulation although a kidney so poisoned produces renin when the renal artery is clamped. Corcoran and Page, and Kohlstadt and Page have come to the conclusion that *reduction in pulse pressure* in the renal vessels is the adequate stimulus. They based their belief mainly upon the following experiment. The artery of a perfused kidney was constricted, while normal blood flow and mean blood pressure in the vessels distal to the constriction were maintained. A reduction in pulse pressure in the renal artery distal to the constricting clamp was followed by the appearance of renin in the renal vein. Reduction in renal blood flow occurred only as a late effect.

But there have been several objections to this explanation. The question has been illuminated by the studies of Trueta and his colleagues upon the renal circulation. It is highly probable that the renal cortex is rendered ischemic by short circuiting of blood (p. 443). This could occur without any appreciable change taking place in the arterio-venous oxygen difference or in the volume of the renal blood flow as a whole.

INHIBITORY PRINCIPLE IN HEALTHY RENAL TISSUE—HYPERTENSINASE

It has been mentioned that ischemia of one kidney, without interference with the kidney of the opposite side is not followed by permanent hypertension nor is the elevation of blood pressure, even while it lasts, as great if a normal kidney is present, and the return of the blood pressure to normal after excision of an ischemic kidney is much less rapid if the opposite kidney is also removed. Also, in order to demonstrate the maximum effect of the pressor substance liberated by the ischemic

kidney the recipient animal must first be nephrectomized. All these facts suggested that normal renal tissue elaborated a substance inhibiting the action of the pressor agent.

Braun-Menendez and his associates demonstrated the presence in normal kidney tissue of an enzyme which inactivates the pressor principle (hypertensin or angiotonin). This enzyme, called *hypertensinase* by its discoverers, and angiotonase by Page and his colleagues, is possibly responsible for the effect exerted by normal kidney tissue upon the development and severity of renal hypertension. Hypertensinase is also found in intestinal mucosa, pancreas, as well as in the erythrocytes, minimal amounts are present in liver and in serum. Its optimum pH lies between 7.5 and 8.5. Attempts to obtain a material from extracts of normal kidney tissue which would neutralize or inactivate the pressor substance responsible for clinical hypertension have not, so far, been very successful.

An antirenin has been produced by injections of purified hog renin into dogs, and into human subjects. Unfortunately, the antirenin obtained in these immunological experiments is not effective in lowering the blood pressure in hypertensive patients. The antirenin so produced is active only against animal renin, not against human renin, it lowers the blood pressure of dogs with experimental hypertension, as well as the blood pressure of dogs with spontaneously occurring hypertension—a condition occasionally found and which is analogous to essential hypertension in man. Human renin is not antigenic and in order to produce an antirenin effective against human renin, it would be necessary to make it so. Experiments are being carried out along these lines. But such experiments could hardly be expected to yield results of much practical value owing to the difficulty of obtaining the large amounts of purified human renin which would be required. Yet if an antirenin produced in this way were effective in essential hypertension, such a result would go far toward proving the humoral nature of the disease.

THE ENDOCRINES IN RELATION TO EXPERIMENTAL HYPERTENSION. Though the evidence points to hypertensin as acting directly upon the vessels, the possibility must be considered that it exerts its effect by stimulating a ductless gland, such as the pituitary or adrenal, to secrete a vaso-constrictor substance. That the adrenal medulla plays a rôle in this respect can be summarily dismissed, for the hypertension is not prevented or modified in any way by bilateral excision of all

medullary tissue. On the other hand, Goldblatt has found that constriction of the renal arteries fails to cause a rise in blood pressure after bilateral removal of the adrenal cortex, even though the animal is maintained in good condition by a high salt and low potassium diet (ch 59). But if cortin is administered to the adrenalectomized animals, renal ischemia is followed by the usual hypertensive response. The blood pressure of hypertensive rats is lowered by adrenalectomy and is only partially restored by the administration of *desoxycorticosterone*. However, these results should not be taken to imply that the renal principle mediates its action *through* the adrenal cortex, for Houssay and his colleagues found that in short term experiments in which an ischemic kidney was transplanted into a non-hypertensive animal, complete adrenalectomy of the latter or ligation of the adrenal veins did not prevent the rise in blood pressure. These two sets of experimental results have been reconciled by the observation that hypertensinogen is reduced after adrenalectomy. The cortical hormone appears to be necessary, therefore, for the production of the substrate upon which renin acts. But desoxycorticosterone is capable itself of inducing hypertension in normal animals, provided that the diet contains adequate amounts of salt. It is ineffective upon a salt poor diet. Damage to the renal vessels (nephrosclerosis) is caused by the administration of this hormone, it also increases the retention of salt (enhanced tubular reabsorption) and an increase in the volume of the extracellular fluids. Both the renal effect and that upon water and salt metabolism are probably responsible for the hypertension caused by this cortical hormone.

There is no evidence that the posterior lobe of the pituitary plays a rôle in the development of renal hypertension, but the anterior lobe appears to be implicated through its action upon the adrenal cortex. Ablation of the anterior lobe in animals with renal hypertension lowers the blood pressure, which is raised to its previous level by the administration of adrenocorticotrophin or of a crude anterior lobe extract. So far as is known neither the thyroid nor the sex glands are in any way responsible for renal hypertension.

THE SIGNIFICANCE OF THE RESULTS OF RENAL ISCHEMIA IN EXPERIMENTAL ANIMALS TO THE HYPERTENSION OF CHRONIC RENAL DISEASE. The immense importance of the experimental work which has been outlined in the foregoing sections and its bearing upon the hypertension secondary to

renal disease requires no emphasis. Direct and conclusive proof that interference with the renal circulation is the cause of the clinical condition and that renin is the responsible pressor agent has not been obtained. But the hypertension of renal disease and experimental hypertension are so similar that it is difficult to believe that both are not brought about by identical mechanisms. It is easy to understand how vascular lesions within the diseased kidney itself could cause a change in renal hemodynamics and the release of renin just as does compression of the renal artery. Further support for such a belief is provided by the finding of renin in the blood of patients with recently developed hypertension⁴. Several clinical observations can be cited which stress the cogent argument offered by the experimental results in explaining the hypertension of renal disease. For example, in unilateral renal lesions, e.g., pyelonephritis, associated with hypertension, removal of the diseased organ is often followed by the rapid return of the blood pressure to normal levels.

The experiments also carry important physiological implications or at least bring up some interesting questions. Does the pressor substance serve as a physiological mediator of renal blood flow? Is it in the nature of a humor or hormone liberated from the renal cells or produced by them when the functional demands upon the kidney call for a greater blood flow through its vessels? A rôle of this nature is suggested by an experiment performed by Drury. A loop of silk thread slightly larger than the renal artery was tied around the vessel of one side in young, growing rabbits. As the animals grew the renal artery increased in size until it reached the diameter of the loop which from then on prevented the renal blood flow from increasing to meet the demands of the growing body. The opposite kidney made up for the deficiency by enlarging enormously and as long as it remained undisturbed the blood pressure did not rise. But removal of the hypertrophied kidney was followed by hypertension (up to 200 mm Hg). An observation reported by Dill and Erickson points in the same direction, namely, that the vasopressor substance serves to proportion the blood flow through the kidney to the demands made upon renal function. The renal arteries on both sides were moderately compressed in a series of pregnant dogs. The operation was followed in

⁴ Renin has not been demonstrated in the blood in hypertension which has existed for some time.

from 48 to 120 hours by hypertension, nitrogen retention and hematuria. The animals had convulsive seizures and died in from 3 to 15 days in coma. Lesions in the liver, e.g., periportal necrosis, similar to those found in eclampsia were found at autopsy. Non-pregnant dogs whose kidneys were subjected to the same degree of ischemia, though they developed hypertension, remained in good health.

OTHER POSSIBLE FACTORS IN THE GENESIS OF RENAL HYPERTENSION The more research that is devoted to experimental renal hypertension the more intricate the problem appears to become, and the greater is the insistence that not one but several interrelated mechanisms are probably involved. Page and Corcoran propose what they call a *mosaic theory* in which nervous and endocrine factors as well as the renin-angiotonin (hypertensin) system play a part.

THE QUESTION OF A NERVOUS FACTOR IN CHRONICALLY HYPERTENSIVE ANIMALS For a time it seemed settled decisively that the hypertension following constriction of the renal arteries was due *entirely* to a humoral mechanism. But the work of Dock, of E. W. Page and Ogden, and of a number of other investigators has thrown doubt upon the purely chemical nature of experimental hypertension throughout all stages of its course. Byron and Wilson, for example, found that removal of a single ischemic kidney from hypertensive rats did not lower the blood pressure. Others have shown that even bilateral nephrectomy did not restore the normal blood pressure of *chronically* hypertensive animals. Evidence pointing in the same direction has been obtained by Dock. He has demonstrated that, whereas pithing a normal rabbit made hypertensive by a continuous infusion of renin does not cause a reduction in blood pressure, the same operation upon a rabbit rendered permanently hypertensive by renal ischemia caused the blood pressure to fall to the normal level. Again, a sympatholytic drug, such as yohimbin hydrochloride, does not ameliorate experimental renal hypertension shortly after its establishment, but has such an effect in hypertension of long standing. As mentioned on page 158 renin has been found in the blood in the acute phase of hypertension, but not in the chronic phase. The sum of these several observations suggests that in hypertension of recent standing a humoral mechanism (renin hypertensin) is responsible, but that the chronic stage is neurogenic. It may be, however, that (1) hypertensin is the

cause of the vasoconstriction in both phases but that in the condition of long standing the vessels are hypersensitive to the pressor principle and respond to very minute, undetectable amounts, (2) the discharge of renin ceases when the blood pressure has risen sufficiently to supply adequately the ischemic kidney, or (3) renin is taken up by the vascular wall and produces hypertensin in this situation rather than in the blood.

THE VEM-VDM SYSTEM Shorr and his associates suggest that hepatorenal factors enter into the mechanism of renal hypertension. They have found that constriction of the renal artery by a clamp causes a derangement of the normal VEM-VDM balance. The initial anoxia produced in the kidney by the constricting clamp is accompanied by the appearance in the blood of the renal vaso-excitatory material (VEM) which normally cannot be detected. A little later the hepatic vasodepressor material (VDM) is found in low concentration in the blood, it increases progressively until in the chronic hypertensive phase it counteracts the action of VEM (ch 27).

PRIMARY, BENIGN OR ESSENTIAL HYPERTENSION (HYPERPIESIA)

The immediate cause of the raised pressure in this as in the preceding type is an increase in the peripheral resistance, i.e., vasoconstriction or narrowing in some way of the peripheral vessels. It is not due to an increase in any of the other factors upon which the maintenance of the normal blood pressure depends, namely, the output of the heart, the viscosity^{*} or the volume of the blood. The calibers of the retinal arterioles are greatly narrowed, appearing upon ophthalmoscopic examination like silver wires. The state of these vessels may be taken as representative of that of the cerebral vessels and probably of the state of the peripheral vessels generally throughout the body.

The systolic and diastolic pressures are elevated equally, or the systolic to a greater degree than the diastolic, 250 mm. is not an unusual figure for the former and 130 mm. for the latter. In other cases, usually those in an advanced stage, the diastolic pressure is proportionately greater than the systolic, and the pulse pressure reduced. The pressure in the pulmonary circuit is believed to be normal. Pulmonary hypertension does, how-

^{*}A moderate increase in blood viscosity has been demonstrated in hypertension but it is a contributory cause only.

ever, occur, as in heart disease (e.g., mitral stenosis) and in certain diseases of the lungs, quite independently of systemic hypertension, and may be present alone or coincidentally with the latter (see also ch 28). The pressure of the cerebrospinal fluid is increased in proportion to the rise in the diastolic pressure.

The pressure in the capillaries and small veins is within normal limits, and the slope of pressure through these vessels is not materially different from that in health. But it follows that the fall in pressure through the arterioles is much greater than in health (fig 16.2). Ellis and Weiss found mean pressures of 155 and 12 mm respectively in the brachial artery and in the capillaries—a fall of over 140 mm. Of this about 125 mm must have occurred in the arterioles. Normally the fall of pressure in the latter vessels is less than half this figure (approximately 50 or 60 mm).

The high pressure in the arterial system, which of course must be overcome by the left ventricle before it can expel its contents, increases the work of the heart by from 40 to 50 per cent. The ventricular cavity becomes dilated and its walls hypertrophied. The cardiac output, except when the circulation is failing, is normal or only slightly reduced. In certain instances, according to some observers, it is increased. The vascular system, including the renal arterioles, in time suffers degenerative changes (p 164) and renal insufficiency may then result (ch 36), a small contracted kidney is usually found after death. It should be emphasized that (as first pointed out by Albutt) the hypertension *develops* in the absence of clinically demonstrable kidney disease. The latter when it occurs is the result of the hypertension. In other words, kidney disease is not a forerunner of essential hypertension but a sequel to it.

PATHOGENESIS The cause of the *increased peripheral resistance* is unknown, though many theories have been proposed.

The pressure is lowered temporarily by drugs which cause vasodilatation, such as histamine, in a number of hypertensive cases examined by Ellis and Weiss the arteriolar resistance was, as in normal persons, removed by this drug. These facts indicate that the vasoconstriction is largely due to spasm (fig 16.2).

Nervous influences It has been suggested that the essential cause of the increased peripheral resistance is an inherent hypersensitivity of the vasoconstrictor nerves and an exaggeration of the

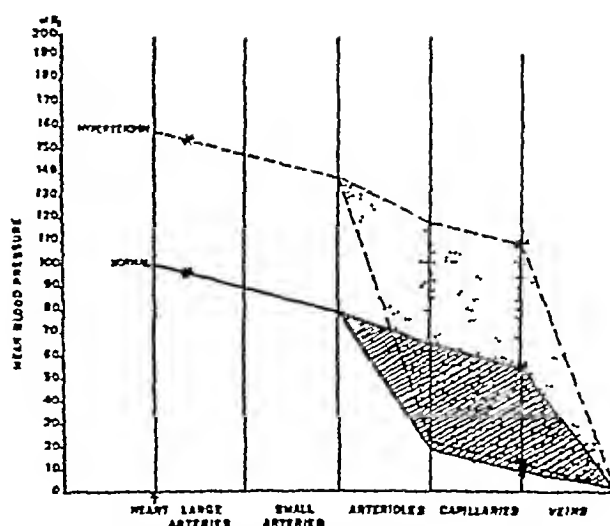


FIG 16.2 Diagrammatic representation of fall in blood pressure in vascular circuit in subjects with hypertension compared with the normal. Shaded and stippled areas represent alteration in the pressure relationships in the skin vessels after the injection of histamine. The lower boundaries of these areas represent the pressure gradient under natural conditions, the upper boundaries this gradient after the injection of histamine (after Ellis and Weiss, modified).

usual vasomotor responses. The demonstration of an hereditary susceptibility to hypertension which is said to be a *dominant* characteristic and transmitted according to Mendelian laws has been pointed to in support of the latter view. Also Lord Dawson in the examination of 650 school children found the blood pressure well above the normal in 8 per cent. Hypertension has also been reported in infants.

In hypertensive subjects the reactivity of the peripheral vessels to nervous stress or to cold, as by the immersion of a hand in ice-water, is usually greater than normal. The latter procedure—the so-called *cold pressor test*—is employed for investigating the vascular responses. In most hypertensives the reaction is excessive, a sharp and inordinate rise in diastolic pressure occurs without an increase in cardiac output or pulse rate, which is taken as indicating a hypersensitive vasoconstrictor mechanism. Hines has studied this reaction in normal as well as in hypertensive subjects over a period of years, and has classed them into three groups, *hyper-reactors*, who respond to the test by a rise over 20 mm Hg, *normoreactors*, who give a rise of between 10 and 20 mm Hg, and *hyporeactors*, with whom the rise was less than 10 mm Hg. The subjects with normal blood pressures and who were *hyper-reactors*, showed a definite susceptibility

to the development of hypertension in later life (see table) In these, as well as in the hypertensive hyper-reactors, the magnitude of the response tends to increase progressively after 40 years of age The afferent limb of the cold pressor response is constituted of fibers mediating the sensations of cold and of pain The efferent impulses pass to the vessels via sympathetic paths The response is abolished by tetraethylammonium chloride, a sympatholytic agent.

The Reaction of the Diastolic Pressure to the Cold Pressor Test, Correlated with the Subsequent Development of Hypertension (Hines)

CLASSIFICATION*	CASES	HYPERTENSION 15 YEARS LATER	
		Cases	Per cent
Hyporeactors	36	0	0
Normoreactors	48	8	17
Hyper reactors	57	31	54

* Blood pressure normal, less than 100 mm Hg diastolic

Though hypertension, temporary at least, may be produced in animals by denervation of the carotid sinuses (p 282) and section of the aortic nerves, there is no evidence that these mechanisms are directly concerned in the production of hypertension in man The response of the carotid sinus to manual pressure (depressor response) is often actually more active in subjects of hypertension than in normal persons Furthermore, in a study made by Thomas of the hypertensive effect of sectioning the sinus and aortic nerves in dogs, it was found that hypertension caused in this way differs from essential or renal hypertension in that increased cardiac output is an important factor in its production, it is cured by complete sympathectomy

Hypertension has also been produced in animals by subjecting them (rats) to loud noises, by stimulation of the hypothalamus, or of certain areas of the cerebral cortex, by the injection of kaolin into the cisterna magna, by ligation of the carotid arteries or by stimulation of the renal nerves But even some of these apparently purely neurogenic types of hypertension appear to be, in part at least, of renal origin For example, the hypertension caused by section of the sinus and aortic nerves is only ameliorated by removal of the sympathetic chains without denervation of the kidney The

latter operation performed subsequently restores the blood pressure to normal It has also been found that renal denervation corrects the hypertension caused by cisternal injection of kaolin

The nervous and emotional strain of modern civilization has frequently been suggested as the underlying factor in the development of essential hypertension There is little evidence which can be cited in support of the idea that mental strain, in the great majority of cases at any rate, is an important factor In an analysis of some 30,000 applicants for life insurance Weiss found the average blood pressure no higher in medical men, teachers, lawyers, business men and others than in farmers who undoubtedly lead a life of greater mental calm It should be mentioned, however, in this connection that Armstrong found in Air Force applicants a high correlation between hypertension and emotional instability, as determined by psychological examination This finding applies, of course, only to healthy young men and not necessarily to women or to males of other ages

Atherosclerosis and arteriosclerosis Arterial degeneration occurs in two main forms

(a) In the first and most serious type of arterial disease plaques of lipid material, cholesterol, cholesterol esters, phospholipids, fatty acids and lipoproteins, appear in the previously thickened intima of the large arteries and their main branches This type is known as *atherosclerosis* (*G atherē*, porridge + *sklēros*, a hardening), *atheroma* or *atheromatosis* The atheromatous areas may later be the seat of calcareous deposits The aorta and the coronary and cerebral arteries are most commonly affected The clinical importance of atherosclerosis lies in its tendency to cause a vascular accident—thrombosis or weakening of the arterial wall and rupture It is now very widely accepted that atherosclerosis is due to some disorder of lipid metabolism, especially of cholesterol The relative concentrations in the plasma of the giant lipoprotein molecules of different densities is altered from the normal, and this, rather than the total concentration of plasma cholesterol (free or as esters) is the important factor in the genesis of atherosclerosis In experimental atherosclerosis, which can be induced in rabbits by high cholesterol feeding, and in dogs, if an antithyroid agent such as thiouracil is given as well, a pronounced increase in the plasma lipoproteins of relatively low density (10 to

20 S_f)⁶ occurs, as determined by ultracentrifugation (applied force, 250,000 times that of gravity) This class of lipoprotein particle is in abnormally high concentration in the plasma of patients recently recovered from coronary thrombosis, and in others with evidence of atherosclerosis The concentration increases with age, and is much higher under 40 years of age in males than in females, but this sex difference tends to disappear in the older age groups These observations are in accordance with the incidence of atheromatosis with respect to age and sex

(b) *Senile arteriosclerosis, Monckeberg's sclerosis*

In this form of arterial degeneration, atrophy or necrosis with subsequent deposition of calcium salts occurs in the muscular coat (media) of the smaller arteries The vessels become tortuous, hard, more or less rigid and brittle, a condition which has suggested the term "pipe-stem" arteries

Neither of the diseases just described is the cause of essential hypertension On the other hand, there is no doubt that atherosclerosis, at any rate, may result from hypertension or be hastened in its development Veins, for example, or the vessels of the pulmonary circulation, which have a pressure much lower than that in the systemic arteries, are rarely the seat of atherosclerosis If, on the other hand, a section of a systemic artery is excised and replaced by a length of vein, atheromatous changes are likely to supervene in the wall of the grafted vessel Arterial changes are also not unusual in the pulmonary vessels when, as in mitral stenosis, the blood pressure in the pulmonary circuit is high One of the earliest experiments on the question of high arterial pressure being a cause of atherosclerosis was undertaken by Josué He produced arterial degeneration in rabbits by daily injections of adrenaline Tyramine, as shown by Duff and his associates, has a similar effect The recent researches of Wakerlin and his associates have shown a close correlation between the height of the blood pressure and the severity of atherosclerosis in dogs with renal hypertension and whose plasma cholesterol had been raised by cholesterol feeding combined with thiouracil administration

Dock considers that a blood pressure under 60 mm Hg as obtains in the smallest arteries and in the veins and vessels of the pulmonary circuit, does not encourage the development of atherosclerosis

This is borne out by the fact that, whereas in persons with a normal blood pressure atheroma does not involve the smaller twigs of the arterial tree, in hypertension the process may extend peripherally into the arteriolar region where the blood pressure may be considerably higher than 60 mm Hg According to Dock, the effect of high arterial pressure upon the arterial wall is not so much that it causes injury (wear and tear) but rather that it provides a certain "percolation pressure" which is required to force the large lipid molecules into the cells of the vascular lining A high rate of blood flow is believed to be also an important factor⁷

A discussion of humeral versus nervous factors in the genesis of essential hypertension

Having in mind the experimental results of constriction of the renal artery in causing hypertension and the probability that renal ischemia is the cause of the high blood pressure in renal disease, the question immediately arises, "Is essential hypertension due to the same mechanism?" The hypertension in animals caused by a moderate degree of compression of the renal arteries (p 157), with its relatively mild course and the absence of renal insufficiency, is comparable to "benign" hypertension in man, that caused by more severe compression of the artery, and accompanied by renal failure and widespread vascular damage, resembles closely the malignant form of the human disease (see below) The observations of Moritz and Oldt may be cited in connection with the renal ischemia theory of essential hypertension They reported that in the histological examination of the kidneys of 100 hypertensive cases in which there had been no evidence of renal disease, arteriolar damage consisting of endothelial hyperplasia, medial hypertrophy and hyaline degeneration of the intima was an invariable finding In the same number of kidneys from non-hypertensive cases, arteriolar disease was absent So far as the systemic vessels were concerned, there was no corresponding distinction between hypertensives and non-hypertensives These observations and those of others

⁷ Arteriosclerosis, by rendering the arterial wall abnormally rigid and thus preventing the expansion of the arterial bed during systole, will tend to raise the systolic pressure, the diastolic pressure, for a similar reason, is normal or even subnormal It is also possible, as Starling suggested, that in some rare instances of arterial hypertension, vascular disease involving the vessels of the vasomotor centers in the medulla, with consequent reduction in blood supply, may be a causative factor

⁶ S_f stands for Swedberg's flotation unit, 1 S_f unit is a rate of migration of the particles in the ultracentrifuge of 10^{-13} cm/sec/dyne/gram

upon post-mortem material apparently indicated that either the walls of the kidney vessels are peculiarly susceptible to the hypertensive effect (i.e., that the renal vascular damage is secondary to the hypertension), which is unlikely, or that the arteriolar changes and the renal ischemia, consequent thereto, are the primary causes of high blood pressure. A more recent study by Castleman and Smithwick has failed, however, to support from histological evidence the renal theory of the origin of essential hypertension. They state that the vascular changes in renal tissue removed at operation from hypertensive cases were not significantly greater than those found in the kidneys of persons with normal blood pressures.

Evidence of a more physiological nature that the renin-angiotonin mechanism is concerned in the production of essential hypertension is provided by the action of antirenin in lowering the blood pressure of dogs with spontaneously developed hypertension—a state comparable to essential hypertension in man.

In Goldblatt's opinion *certain* cases of essential hypertension, at any rate, are of humoral origin.

But probably the strongest evidence for essential hypertension being of renal origin has been secured by Goldring and his associates. In a study of 60 patients with essential hypertension of all grades of severity, the tubular excretory mass of the kidney (p. 471) was reduced in all but 3. The fraction of fluid filtered from the plasma was increased (see action of hypertensin, p. 159) due apparently to constriction of the glomerular efferent vessel and, as a consequence of this, a rise in intraglomerular pressure.

The success which has followed the surgical treatment of essential hypertension, in its chronic stages, points to a nervous rather than a humoral mechanism playing the important role. The surgical operations vary in their extent from bilateral removal of segments of the splanchnic nerves (greater and lesser), and of the 9th, 10th, 11th and 12th sympathetic ganglia (Peet), to a complete sympathetic resection on both sides from the 3rd or 4th thoracic ganglion to below the 2nd lumbar (Grimson). Peet and Isberg have reported their results in over 400 patients who had been operated upon from 5 to 12 years previously. Of the total, i.e., those who before sympathectomy were suffering from the disease in all degrees of severity, 57 per cent were then living. Of those in whom the disease was of mild degree and showed neither renal,

cardiac, or cerebral manifestations, 95 per cent survived for from 5 to 12 years, while 19 per cent of those suffering from malignant hypertension before operation were living at the time of the report (1946). Sixty-five per cent of those who before operation had cerebral episodes had no recurrences in the subsequent years. Similarly favorable results of sympathectomy have been reported by others.

The hypotensive action of tetraethyl ammonium chloride, dihydroergocornine and other adrenergic blocking drugs (ch. 59), used in the treatment of hypertension, also supports the conception of a neurogenic factor in essential hypertension of long standing. These drugs have no effect upon the action of hypotensin.^{*} It should also be mentioned that hypertension has been produced in dogs by chronic electrical stimulation of the renal nerves by means of in-dwelling electrodes.

MALIGNANT HYPERTENSION

In this very severe form of arterial hypertension there are widespread degenerative changes with occlusion of the peripheral vessels. The media of the arterioles and small arteries show hypertrophy of their muscular tissue and fibrosis. Hyperplasia and degenerative changes are found in the intima. Small hemorrhages are commonly found in various regions due to the necrosis of the walls of the vessels. As a result of the destruction of the glomeruli, renal failure ensues and death in uremia may occur. More usually death is by cerebral hemorrhage. This condition is probably of the same nature as the benign form but of a much more severe grade.

Eclampsia gravidarum

Eclampsia is a state occurring in pregnancy (usually after the fifth month) characterized by hypertension, headache, nausea and vomiting, albuminuria, edema, tremors, or convulsions ending in coma. The blood pressure may be extreme and of rapid development. The diastolic pressure shows the more pronounced rise, indicating a generalized

^{*} These drugs are used in the treatment of clinical hypertension, or as an aid in the selection of suitable cases for operation. When the systemic vessels are narrowed as a result of organic changes, it is not to be expected that denervation will be of benefit and nerve-blocking drugs will have no effect upon such changes. A distinction between organic narrowing of the vessels and a neurogenic vasoconstrictor factor can also be made by means of caudal anesthesia, which abolishes the latter but exerts no effect upon the structural changes.

vasoconstriction. The cause of the arteriolar spasm is unknown. The blood pressure, as a rule, falls after evacuation of the uterus, which points to the uterine contents being in some way responsible for the symptoms. Since death of the fetus does not result in an abatement of the condition, and it develops in association with a hydatidiform mole, placental tissue is certainly incriminated. The convulsions appear to be due to spasm of the minute vessels of the brain. Beyond these facts nothing is known for certain with respect to the causative mechanisms in eclampsia. A toxic substance formed by the placenta may exert a direct constrictor effect upon the systemic vessels, or on the vessels of the kidney, in the latter instance the immediate cause of the hypertension being of renal origin. Again the placental material might act through an endocrine. Some (Anselmino and Hoffmann) have attributed the convulsions to the antidiuretic principle of the pituitary and its effect in causing water retention. Theobald has obtained an antidiuretic material from eclamptic cases, though he denies that it is of pituitary origin. But there is no substantial evidence that the pituitary or any other endocrine plays a part in the development of eclampsia. Attempts to demonstrate a pressor principle or other material in placental extracts which could account for the symptoms have failed. It is possible that the toxic substance, whatever it may be, is produced by placental ischemia, infarcts, larger and more numerous than normally are found in the placenta in most cases of eclampsia.⁹

HYPOTENSION

In an adult, a systolic pressure which is persistently below 110 mm Hg and for which no cause can be found is referred to as essential or primary hypotension. The subjects of *essential* hypotension, beyond showing possibly a greater susceptibility to fatigue, suffer no ill effects, on the contrary they are more likely to be free from cardiac and renal disease, the condition for this reason being said to forecast longevity. Robinson states "Hypotension is not a disease, it is the ideal blood pressure level."

Apart from the foregoing type, low blood pres-

⁹ Evidence for such a mechanism has been obtained by Ogden and his colleagues. Reducing the blood supply to the pregnant uterus of dogs (in acute experiments) by partially occluding the aorta below the renal arteries was followed by a rise in blood pressure. This procedure has no such effect in non-pregnant animals.

sure occurs either as a temporary or a persistent phenomenon in many conditions. Some of these are *hemorrhage, traumatic shock, anesthesia, tuberculosis* and *debilitating diseases* of various kinds. It may occur also as a result of the vasodilation and cardiac asthenia associated with *acute fevers*, from *myocardial failure*, and in *Addison's disease* or in *hypothyroidism*.

Orthostatic or postural hypotension is an unusual but interesting condition in which the reflex mechanisms normally operating to maintain the blood pressure against the effect of gravity are apparently in abeyance, or their sensitivity greatly depressed (p 169). A profound fall in blood pressure occurs in the standing position, the systolic pressure may fall to 40 mm Hg and the diastolic to zero (as determined by the usual method). The subject experiences dizziness or may fall in a faint. Other features of this condition are (a) absence of sweating, (b) failure of the pulse rate to increase upon rising from recumbency to the standing position, (c) slight depression of the basal metabolic rate, (d) loss of sexual desire, (e) skin pallor and (f) blood urea around the upper limit of normal. The inability to sweat, taken together with the failure of the reflex vasoconstriction to a change in posture, suggests some fundamental abnormality of the sympathetic nervous system. The administration of benzedrine sulphate effects some temporary relief by reducing the extent of the fall in blood pressure in the standing position.

Induced or controlled hypotension. Drugs of the methonium series—pentamethonium bromide or iodide and the corresponding hexa- or decamethonium salts—which block tonic sympathetic vasomotor impulses and lower the normal blood pressure have come into use to reduce blood loss in major surgical operations.

VENOUS BLOOD FLOW

FACTORS WHICH INFLUENCE THE FLOW OF BLOOD IN THE VEINS

It has been mentioned (p 143) that the blood pressure slopes gradually through the systemic capillaries and veins to the right auricle, but that with the narrowing of the venous bed from the "capillary lake" to the right side of the heart the velocity of flow increases. The following are the factors which influence the venous flow.

- (1) Contraction of the left ventricle
- (2) The quantity of blood flowing through the

arterioles from the arteries in relation to the capacity of the capillaries and veins

(3) The subatmospheric pressure within the thorax

(4) The action of the right side of the heart

(5) The massaging effect of the skeletal muscles and the support afforded by the abdominal wall

(6) The effect of gravity

(1) *The contraction of the left ventricle* (the "vis a tergo") The energy of the ventricular contraction as we have seen is expended in driving the blood through the arterioles and onward through the capillaries and veins to the right auricle. By the time the blood reaches the auricle the energy has been almost entirely dissipated in overcoming the frictional resistance offered by the vascular channels, at this point the pressure is only about 5 mm of water

(2) *The quantity of blood flowing through the arterioles in relation to the capacity of the capillaries and veins* Generally speaking, the more blood which is received from the arterial side the greater will be the capillary and venous pressures. That is, with dilated arterioles the difference between the arterial pressure on the one hand and capillary and venous pressures on the other, tends to be reduced, with constricted vessels the pressure difference will be increased. The relationship, however, does not always hold for the reason that the capillaries and veins are capable of adjusting their capacities to the quantity of blood received from the arterial system. The minute vessels may dilate to accommodate the extra blood, so that little change in capillary or of venous pressure will result. Bayliss and Starling, for example, found that when the arterioles were opened as a result of the stimulation of a vasodilator nerve, the expected rise in venous pressure did not always occur, and Hooker observed that when the hands were warmed, more blood entered the veins but a rise in venous pressure did not result. On the other hand, if the arterioles constrict the more peripheral vessels may reduce their capacities and thus maintain the pressure in spite of the reduced amount of blood which they receive

(3) *Subatmospheric pressure within the thorax*

The intra thoracic pressure has an important influence upon the flow of blood along the great veins of the thorax and abdomen and consequently upon the filling of the right auricle. During inspiration the pressure within the thorax is about -6 mm of mercury (81 mm H₂O) below that of the atmos-

phere. During expiration it amounts to about -2.5 mm of mercury (34 mm H₂O). The subatmospheric pressure expands the thin-walled intrathoracic veins and the venous blood is sucked into the thorax. A similar effect but of less degree is exerted upon the walls of the auricles. The diameters of the thick-walled ventricles, however, and the comparatively rigid coats of the larger arteries remain practically uninfluenced by the "negative pressure" during ordinary breathing. It has been mentioned that the blood in the great veins at their entrance into the auricle has a small but definite pressure of about 5 mm H₂O transmitted from the arterial side. That is to say, if the thorax were opened so as to abolish the subatmospheric pressure within it and a manometer placed in the inferior vena cava a positive pressure of this magnitude would be registered.

The flow of blood toward the heart is thus furthered by both the positive pressure in the venous system and the suction pressure exerted by the thorax. The sum of these two is sometimes spoken of as the *effective venous pressure*. Thus, if the positive pressure in the great veins at the right auricle amounts to $+5$ mm H₂O and the pressure within the thorax to -80 mm, then 85 mm H₂O represents the effective venous pressure. Since the negative pressure in the thorax increases during inspiration, the effective venous pressure must likewise increase and auricular filling be hastened during this phase of respiration. The descent of the diaphragm during the inspiratory phase, also, by compressing the abdominal contents, increases the pressure in the inferior vena cava (the femoral veins and sometimes the iliacs being provided with valves) and augments the flow of blood toward the heart. The thorax thus acts as a pump which "lifts" the blood as well as "forces" it toward the heart (see fig. 29.3, p. 349). Sometimes when vigorous respiratory efforts are made, slight fluctuations of the venous pressure can be detected in the peripheral veins of the human subject. These variations rarely amount to more than 10 mm of H₂O but may be considerably higher than this when dyspnea resulting from obstruction to the free entrance and egress of air from the lungs exists. Owing to the inertia of the blood column the aspirating effect is less evident the nearer to the periphery at which the pressure measurements are made.

The respiratory effects upon venous pressure can also be exaggerated in normal subjects by the following procedures. If a forced expiration is

made with the glottis closed (Valsalva's experiment) the negative intrathoracic pressure can be abolished and a positive pressure of several millimeters of mercury substituted. The veins of the neck, face and limbs become distended with blood as a result of the impediment to the flow into the right auricle. The peripheral venous pressure under these circumstances may rise to 400 mm H₂O or more. In the converse experiment of Mueller in which a forced inspiration is made with the glottis closed, the powerful suction effect may cause a fall of 50 mm H₂O or so in the venous pressure of a peripheral vein. The increase in negative intrathoracic pressure may be seen by means of the X-ray to exert an effect upon the ventricle which during diastole becomes somewhat enlarged beyond its usual size.

(4) *The action of the right side of the heart upon the blood-flow in the veins.* Obviously if the blood is not passed on again by the right heart as quickly as it is carried to it by the great veins, the velocity of the blood-flow in the venous system will be reduced. As a result the venous pressure will rise. That is, there will be a tendency for the blood to be "dammed back." When the heart is beating vigorously and output and inflow are balanced no rise in venous pressure occurs. In health the force of the ventricular contraction is nicely adjusted (p. 246) to the quantity of blood which pours into the auricle from the veins and no accumulation occurs. If, however, the heart fails the venous pressure rises and back pressure effects ensue.

The influence of cardiac action upon the movement of the blood in the veins must not be taken to imply that the heart exerts any aspirating effect. The ventricle does not "draw" blood from the auricle and great veins when it relaxes and expands as one would suck up fluid by means of a bulb-syringe. Such an action has been suggested but it is difficult to conceive how any significant effect of this nature could result even were a negative pressure appreciably below that within the thorax created within the heart chambers. The veins are thin-walled and any reduction of pressure upon their outer surfaces, as we have seen, causes them to expand. But a lowering of pressure in their interiors would cause them to collapse and the venous flow would be blocked automatically, or at least reduced, before the inertia of the blood had been overcome. Moreover, intraventricular pressure curves fail to show a negative pressure during cardiac relaxation (fig. 21.3) as might be expected if the ventricle exerted a suction effect upon the blood in the great veins.

It may be mentioned, however, that during *systole*

the ventricle does cause a slight fall in pressure in the auricle and in the great veins due to the drawing down of the floor of the auricle (p. 206). Also as a result of the ejection of 60 cc. or so of blood from the thorax at each beat a slight but sharp increase in the negative pressure within the thorax is induced during the cardiac contraction. This is reflected in the veins, auricles and other thin-walled intrathoracic structures. It may even be detected within the thoracic portion of the esophagus by placing a balloon therein and recording the pressure change. These variations in intrathoracic pressure occurring during the contraction of the ventricle are known as the negative heart pulse or the cardiopneumatic movements.

(5) *The massaging effect of the muscles.* The intermittent pressure which is brought to bear upon the blood by contraction of the limb muscles aids in propelling it towards the heart. The arrangement of the valves of the veins serves to give the blood flow this direction. The muscles thus act as subsidiary pumps which aid very materially the flow of venous blood, especially during muscular exercise, but to a lesser extent at all times, except when there is complete muscular relaxation. In strenuous exercise, on account of the much greater amount of blood entering the veins from the arterial side, there is a tendency for the venous pressure to rise, but this to a large extent is compensated for, provided the cardiac action is unimpaired, by the increased aspirating effect of the respiratory movements. The muscles of the abdominal wall also contract during exercise to lend support to the abdominal veins and prevent them from becoming over-distended. The venous reservoir is in this way not permitted to enlarge its capacity unduly for the accommodation of the increased volume of blood which, in consequence, is borne onwards to the right auricle.

(6) *The effect of gravity.* Besides the hydrodynamic factor, i.e., the energy of the cardiac contraction, in the development of pressure within the vascular system, there must also be considered the hydrostatic factor, or weight of the blood column, which comes into play when the erect posture is assumed. It is convenient to consider the arterial and venous systems together in their relation to the hydrostatic effect. Above the level of the heart gravity opposes the hydrodynamic factor in the arteries but aids it in the veins. Below heart level the reverse is true, the hydrostatic and hydrodynamic factors being summed in the arteries but opposed to one another in the veins.

In the case of the *arteries* the effect of gravity

upon the blood in the vessels above heart level is fully compensated. That is, the pressure in the brachial artery is as high or actually higher when the subject is standing than when he is lying down. The pressures in the brachial and tibial arteries are approximately equal in the latter position. Unlike the brachial pressure, the pressure in the arteries of the lower limb varies widely with the position of the body. When the subject is in the vertical head down position or in the L position (lying on back with lower limbs vertical) the pressure in the artery of the leg is much lower than it is in the recumbent position. In the standing position, on the other hand, the pressure in the leg artery is higher than in the lying down position (lower limbs horizontal), owing to the hydrostatic effect, the nearer the foot that the measurement is made the higher will be the observed pressure. The difference between the pressures in the brachial and the artery of the leg in these different positions is equal to the height of a blood column which would reach from one artery level to the other. In the case of the blood supply to the upper part of the body nervous mechanisms (ch 27) governing the caliber of the arterioles (especially of the splanchnic area) are called into play to antagonize the gravity effect. An adequate pressure in the cerebral vessels is thereby assured, but there is no corresponding mechanism for the maintenance of a constant blood pressure to the lower limbs, nor, in the natural positions of the body, is it required. The pressures in the brachial and posterior tibial arteries in different positions of the body are given in table 17.

Though the systolic pressures in the femoral and brachial arteries are usually about equal in the horizontal position, a marked difference between the two is seen in aortic regurgitation, the femoral systolic pressure averaging some 50 mm Hg higher

than the brachial pressure (L. Hill and Rowlands). A "differential" pressure of this character is also observed in toxic goiter (average 37 mm Hg higher in femoral) and in arteriosclerosis (average 8 mm higher in femoral). It may also be seen in normal persons after muscular exercise. The differential pressure is attributed by Bazett to the greater resistance offered to the flow of blood in the femoral artery and the conversion, in consequence, of a greater proportion of kinetic energy into stress as the blood stream is slowed. The greater mass of the blood column entering the femoral artery as compared with that entering the brachial is also a factor.

It is through the *venous system* that the effects of gravity upon the circulation are the more prominently displayed. This is on account of the lower venous pressure, the greater distensibility of the venous walls, and also of the fact that the height of the blood column which must be raised against gravity is much greater (from feet to heart) than that of the arterial blood column (from heart to brain). In man the mechanisms whereby the effects of gravity upon the venous system are offset are remarkably efficient. In monkeys also, the compensatory devices are well developed and in some varieties of these and in the anthropoid apes the effect of gravity is counteracted as effectively as in the human subject. In many of the lower animals, on the contrary, there is little evidence of compensatory mechanisms.

The factors which enable the blood in the veins below the thorax to overcome the gravity effect and to be carried to the level of the auricle are several. Let us consider for a moment a U-shaped tube with rigid walls. If liquid be permitted to flow into one limb of such a tube, the liquid, being supported by the rigid walls, rises in the other limb until it overflows. After the liquid has been al-

TABLE 17
Showing effect of gravity upon the arterial blood pressure
(Modified from Hill and Flack.)

POSTURE	BRACHIAL ARTERY PRESSURE	POSTERIOR TIBIAL PRESSURE	PRESSURE DIFFERENCES OBSERVED	PRESSURE DIFFERENCES CALCULATED FROM HEIGHTS OF BLOOD COLUMNS SEPARATING POINTS IN TWO ARTERIES
	mm Hg	mm Hg	mm Hg	mm Hg
Horizontal	106	106	0	0
Standing	110	165	55	58
Vertical, head down	115	50	65	63
L position, legs up	115	85	30	33

lowed to come to rest, then at any level of the tubing the hydrostatic pressure corresponds to the height of the liquid column extending above that level. The aorta with the main arteries of the lower limb, and the inferior vena cava with its tributaries, are roughly comparable to two limbs of a U-tube. But the venous walls are highly distensible and quite incapable of supporting the blood column, and, unless the gravity effects were overcome, the blood would sink and "find its own level." Furthermore, a system of small vessels—capillaries and venules—of variable capacity is interposed between the limbs of the U. In the corpse, for example, the blood subsides to the dependent parts, and the hydrostatic pressure when the body is placed erect amounts to no more than a few centimeters of water.

Though the factors which aid the venous flow have already been dealt with, those especially concerned with raising the blood against gravity may be given again here. They are —

(a) The impetus given to the blood by the left ventricular contraction (vis a tergo)

(b) The abdominal and limb muscles support the vein walls and prevent their "giving" under the weight of blood. By this means the veins in a sense come to simulate rigid tubes. When the abdominal muscles are weakened or paralyzed the support which they normally provide is seriously impaired, but may to a large extent be restored by a tight abdominal bandage. The intermittent contractions of the skeletal muscles in conjunction with the valves of the veins propel the blood in the upward direction. The important part played by the muscles in the return of blood to the heart is shown by experiments in which the body is passively tilted into the upright position. Persons who fail to show a rise in intramuscular pressure when tilted passively are likely to faint.

(c) The suction and force-pump action of the respiratory movements

(d) The veno-pressor and capillary tonus mechanisms which, through the activity of the sinus and aortic nerves (p. 282), control the calibers of the small veno-capillary vessels of the splanchnic bed and prevent pooling of blood in this area.

If any or several of these factors fail, accumulation of blood (venous stasis) in the dependent parts of the body is likely to result. Thus, in the human subject if, upon assuming the erect posture after a protracted confinement to bed, the muscles of the abdomen and limbs are weak and the tone

of the nervous mechanism governing the peripheral vessels lowered, the hydrostatic effect is overcome with difficulty. The blood subsides into the capacious abdominal veins and capillaries and the right heart is no longer adequately supplied with blood. The arterial pressure falls and the cerebral blood flow becomes inadequate. The subject turns pale, sweats freely and feels giddy or perhaps nauseated, and in a complete faint loses consciousness (syncope). See fainting, p. 287.

The effect of gravity upon the circulation in an animal such as the dog or cat which normally is fairly well able to compensate for postural changes, and the importance of the various factors comprising the control mechanism have been well shown by L. Hill. The medullary control over the peripheral vessels was removed by section of the cord at the level of the 1st thoracic vertebra. The tone of the abdominal muscles also is lost after this operation and the respirations are altered. The blood pressure falls even when the animal is in the horizontal position. When placed in the feet-down position, the pressure drops to zero. If the animal's abdomen be compressed the capacity of the splanchnic vessels is thereby reduced and blood forced upwards to fill the right heart. The blood pressure is raised again and the circulation restored. In the head-down, feet-up position the rigid cranial wall supports the blood column, the heart fills and the carotid pressure rises. When the main nerve trunks (splanchnics, ch. 72) carrying fibers to the peripheral vessels are divided alone the respiratory pump in part compensates for the gravity effect, and though marked changes in pressure occur when different positions are assumed the circulation is maintained. Edholm found that the blood pressure of normal cats under chloralose anesthesia fell by 34 mm Hg on the average when the animal was held in the vertical feet-down position, but recovered shortly to a final pressure within 30 mm Hg of its previous level in the horizontal position. It appears from Edholm's results that pooling of blood under the effect of gravity occurs in the liver and not to any great extent in the splanchnic vessels generally, for the fall in blood pressure in the feet-down position was about the same after evisceration as before, whereas, after removal of the liver from eviscerated animals, the vertical feet-down position caused a fall in blood pressure of only from 5 to 10 mm Hg. That the splanchnic vascular bed is, nevertheless, a factor in the compensatory mechanism tending to oppose the effect of gravity was indicated by the observation that the final height of the blood pressure in the feet-down position was higher in normal than in eviscerated animals.

The circulation of animals such as the domestic rabbit or the snake, which have not acquired a compensatory mechanism, is placed at a great disadvantage when

the vertical position is assumed.¹⁰ This has been shown very clearly upon the latter species in the following experiment. The heart and abdominal vessels were exposed and the reptile fastened to a board. The right side of the heart was seen to be adequately supplied with blood so long as it was held in the horizontal position. When placed vertically the auricle was no longer properly filled and the great veins entering the heart were nearly empty. When the preparation was then immersed in a cylinder of water up to the heart level the hydrostatic pressure of the venous blood was counterbalanced by that of the surrounding water, and the blood again flowed freely into the heart chambers.

VENOUS PRESSURE

For the *indirect* measurement of the venous pressure in man the instrument devised by Hooker and Eyster is commonly employed. It is based upon the principle first employed by von Recklinghausen.

The instrument consists of a small round chamber with a glass top and metal walls (fig. 16.3). Over the bottom of the chamber is stretched a piece of rubber dam with an opening about half an inch square in its center. Lengths of rubber tubing connect the cavity of the chamber, respectively, with a water manometer and a small hand bulb. The rubber dam is moistened with glycerine and placed with its central opening lying over the vein to be examined. The pressure within the chamber is raised by a few compressions of the bulb. The venous pressure will equal that required to cause collapse of the vein. This pressure is read from the manometer at the instant that collapse occurs. All determinations are made with the vein just below the level of the auricle so as to annul the hydrostatic effect. The part under examination is supported in order to insure complete muscular relaxation and the subject should be at rest for 15 minutes prior to the determination. If the pressure is measured with the subject recumbent, the vein, e.g., median basilic or a vein of the forearm or hand, is placed at a point situated one third of the distance from the sternum to the back at a level of the 4th intercostal space. If the measurement is made in the sitting position the vein is brought to the level of the 4th interspace.

The venous pressure may also be determined *directly* though less conveniently and with more discomfort to the subject (but somewhat more accurately) by the insertion into the vein of a hollow wide bore needle, as first employed by Moritz and von Tabora.

A simple indirect method which gives approximate results is that of Gaertner. With the patient in the supine position the hand is first lowered below heart level until the veins on the back of the hand fill. The

part is then slowly raised until the veins collapse. The distance of the vein in millimeters above the right auricle at which collapse occurs gives the venous pressure in millimeters of blood. The vein represents a manometer tube connected with the auricle and the level at which the vein empties indicates the point where the venous pressure at the auricle just about equals the weight of the blood column. Another approximate but ready method is the inspection of the veins on the under surface of the tongue. These normally are collapsed when the body is standing or in the sitting position, but are distended if the venous pressure is higher than about 200 mm H₂O.

NORMAL VENOUS PRESSURES

The pressure in the median basilic vein taken with the body recumbent shows considerable variations in different individuals. The range is between 60 and 100 mm H₂O. The pressure is greater in veins nearer the periphery and shows a progressive diminution toward the heart. It must be remembered that the pressure *as measured* is the algebraic sum of the positive pressure transmitted from the arterial side and the negative pressure exerted upon the blood column from the thoracic cavity. The effect of the latter is slight in the peripheral veins, but as the thorax is approached

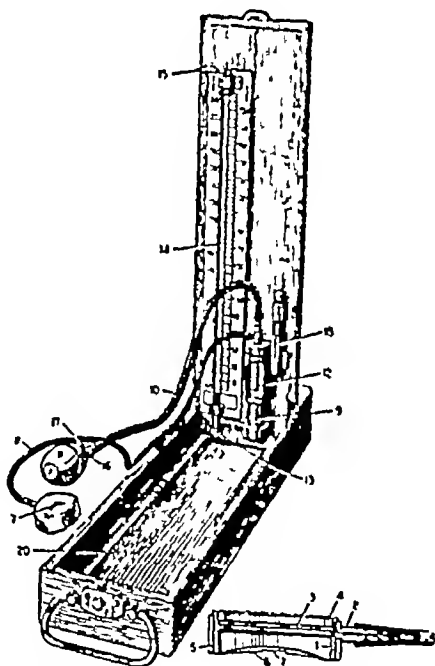


FIG 16.3 Apparatus for the determination of venous blood pressure in man by the indirect method. For detailed description of parts see "Clinical measurement of venous pressure," Eyster Macmillan 1929

¹⁰ Sheep sometimes collapse and die when held up right during shearing

the venous blood comes decidedly under its influence. According to Burton-Opitz the decline in pressure along the peripheral veins amounts to 1 mm Hg (13.5 mm H₂O) for each 35 mm distance. The point of zero pressure, i.e. where negative and positive pressures just balance one another, is in the case of the upper part of the body in the jugular vein at the root of the neck, below this point the pressure is negative.¹¹ As mentioned above, it varies between -35 and -80 mm H₂O at the right auricle.

The point of zero venous pressure below the heart is at about the level of the entrance of the hepatic veins into the vena cava (Burton-Opitz).

The venous pressure varies, of course, as a result of the gravity effect, with the position of the vein in relation to the auricle. In the veins of the hand, for example, the pressure is lower by from 50 to 100 mm H₂O when the hand is at heart level than when it is dependent. The differences in pressure resulting from changes in the vein's level will obviously be larger in the case of the lower limbs. But since certain factors, already mentioned, enter to modify the gravity effect the actual differences which are observed do not coincide with the height of the column of blood from the vein level to the heart. The greater the velocity with which the blood flows upwards the greater will be the discrepancy between the calculated pressure and that as actually measured (see also p. 140). For example, the hydrostatic pressure of a column of blood extending from the foot to the heart would

¹¹ For this reason there is supposed to be danger, should a vein be nicked in this region during an operation, of air being sucked in, and carried to the heart. The air might then be whipped up with the blood and cause frothing within the cardiac chambers, with resulting acute cardiac failure or multiple pulmonary emboli, or even reach the arterial system and cause blockage of a cerebral or a coronary vessel. The fear of air entering the vein under such circumstances, however, would appear to be exaggerated. The walls of the vein tend to collapse under the suction pressure within and automatically close the lumen. Furthermore, even should a small amount of air enter the vein, it is unlikely that serious effects would ensue, since to produce the effects just described in an animal relatively large quantities of air must be injected. Coles, Richardson and Hall have deduced from their experiments upon dogs that about 500 cc of air in the circulation would be lethal for a human being. A more likely portal for the entrance of air into the circulation is via the uterine veins during the induction of abortion or following labor at term. Several fatalities supposedly due to this cause have been reported. It may also occur during the injection of air into the pleural cavity for the production of pneumothorax. It is probable, however, that in many instances in which death has been attributed to air embolism it has been due to some other cause.

amount to over 1200 mm H₂O. The pressure as actually measured in the vein of the foot in the standing position is only a fraction of this. The effect of muscular exercise upon the venous pressure has been considered (p. 169). The relation of venous pressure to heart rate and cardiac output is dealt with on pages 246, 252, and 263.

VARIATIONS IN VENOUS PRESSURE

Muscular exercise causes a prompt rise in the pressure of blood in the veins draining the exercised parts. Normally, a rise of from 20 to 50 mm H₂O occurs during the exercise, but falls within 30 seconds or so after cessation of the work to the original level.

Obstruction of a peripheral vein will, of course, cause a local rise in venous pressure and compression of the large intrathoracic veins will be followed by a general rise.

Congestive heart failure In cardiac failure the pressure in the median basilic vein may reach a

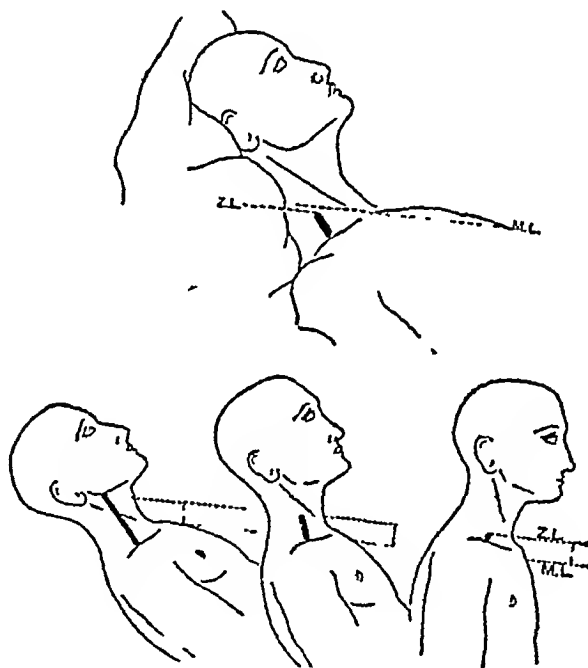


FIG 164 *Upper drawing*, normal subject supine with head on pillows. The zero level (Z.L.) indicated by the top point of swelling of the external jugular vein, lies a little below the manubrial line (M.L.) *Lower drawing*, illustrating excessive venous pressure equivalent to about 8 cm of water. In the upright position the zero level may be just a little above the clavicle, as the subject is inclined the zero level comes to occupy a higher and higher position in the neck until it reaches to the angle of the jaw. Its vertical distance above the manubrium alters very little, however, with changes in position (after Lewis).

value of over 300 mm H₂O. The cause of the high venous pressure has been, and still is, disputed. According to one view, it is due to right ventricular weakness and a consequent "back pressure" effect upon the venous system (fig 164). The work of Warren and Stead (p 38) points rather to an increase in plasma volume as the main causative factor. The increased pressure distends the veins and capillaries of the portal system, the liver becomes enlarged and the other abdominal viscera engorged. The rise of pressure in the systemic veins and capillaries increases the filtration pressure, and is a factor in the production of edema. The slowing of the blood flow through the small vessels of the skin may cause cyanosis. Both edema and cyanosis tend to be more pronounced in the dependent parts (e.g., feet and ankles). The rise in pressure in the cerebral veins leads to slowing of the intracranial circulation, especially when the patient is recumbent and the venous return from the head is, therefore, not aided by gravity. Failure of, or increased resistance on, the left side of the heart causes a rise in pressure in the left auricle which is transmitted to the pulmonary circuit and ultimately may lead to failure of the right ventricle

(see also p 258). The lung vessels distend and encroach upon the air spaces.

In *hypertension*, except when myocardial failure is impending or has supervened, the venous pressure is within normal limits.

In the stage of excitement during the induction of *anesthesia* a very considerable elevation of the venous pressure may occur. The effects of disturbances of the respiratory mechanism upon venous pressure are also seen in *asthmatic attacks* (p 428) and in *Cheyne Stokes breathing* (p 413). The pressure rises to 180 mm H₂O or so during the phases of apnea of the latter condition, but approaches the normal value during the periods of hyperpnea. Reduction in the carbon dioxide content of the blood, if extreme, lowers venous pressure by increasing the tone of the peripheral vessels.

In severe *hemorrhage* and *surgical shock* the venous pressure is subnormal.

Local elevation of venous pressure, e.g., increased pressure in arm veins with normal venous pressure in lower limbs, or vice versa, is seen in obstruction to the venous flow in the superior or in the inferior vena cava, respectively.

CHAPTER 17

THE VELOCITY OF THE BLOOD

The relative blood velocities in different parts of the circulation are dependent upon the sectional areas of the respective vascular beds (see p 145). The velocity of the blood through the circulation as a whole, on the other hand, varies in accordance with the quantity of blood ejected per unit of time from the heart.

The velocity of the blood may be considered from four different aspects which are indicated by the following terms:

- (a) The mean lineal velocity
- (b) The circulation time
- (c) The volume flow
- (d) The circulation rate or output of the heart

The output of the heart (d) will be dealt with later (chapter 26). The other three measurements will be considered now.

THE MEAN LINEAL VELOCITY

By this term is meant the distance which the column of blood travels in a measured period of time along a given vessel, e.g., the carotid, femoral, etc. In very small transparent vessels, for instance those of the mesentery of a mammal or those in the tongue or toe-web of a frog, the blood velocity can be measured by timing the progress of a red cell beneath the microscope. In the large vessels the velocity can be calculated if the quantity of blood which passes a given point in a unit of time and the cross area of the vessel are known. It is, of course, not permissible merely to sever an artery and measure the amount of blood which escapes over a certain period, since such a procedure would completely alter two of the most important factors in the circulation, namely, the total blood volume and the peripheral resistance. Several instruments have been devised for measuring the volume of the blood flow in animals while the normal circulation is maintained.

Of such instruments the *Stromuhr* of Ludwig is one of the oldest and the best-known (fig 17 1). It consists of an ingoing and an outgoing cannula which are inserted, respectively, into the proximal and distal sections of a divided artery. The blood flowing from the artery into the instrument enters a small pear-shaped flask of known capacity filled with oil. Upon the entrance of the blood the oil is forced over into

another flask of identical size but which has previously been filled with saline. The entrance of the oil forces the saline in turn into the peripheral section of the artery. When flask 1 is filled with blood and flask 2 with the transferred oil, the instrument is rotated through a semicircle. This reverses the positions of the flasks and the process of filling and emptying is repeated. The blood in flask 1 is now forced into the peripheral part of the artery as blood enters flask 2 and displaces the oil as before. From the number of fillings of the flasks during the period of observation the volume of blood flowing in a unit of time is readily calculated. This value divided by the cross area of the vessel, since velocity is inversely proportional to the sectional area, gives the lineal velocity. Thus $V = v/\pi r^2$ where V = velocity in millimeters per second, v the volume of blood in cubic millimeters flowing into the instrument in a given time and r the radius of the artery in millimeters. For instance, if the flow is 3000 c mm per second and the diameter of the vessel is 4 mm, the cross area is $\pi (\frac{1}{2})^2 = 3.14 \times 4 = 12.56$ sq mm and the velocity is $3000/12.56 = 238$ mm per second. This determination, of course, ignores the fluctuations in velocity which occur during the cardiac cycle,¹ and in consequence gives only the average or mean velocity. It may also be recalled that the velocity is not the same at different points along the radius of the blood column, the flow being much greater in the axial part of the stream than toward the circumference.

THE *thermo-stromuhr* OF REIN In this method the blood as it flows past a point in the vessel (which is exposed but not opened) is heated by means of electrodes connected to a high frequency current. The temperature of the blood is measured below and above the heated region by means of thermojunctions. The magnitude of the temperature difference is a function of the blood flow and is used as the basis for the calculation of the latter. The temperature difference diminishes with increased blood flow and vice versa. This method is accurate and possesses the advantage that

¹ The velocity rises during systole and falls during diastole. There is therefore a pulse of velocity as well as of pressure. Green and his associates found that the velocity of blood flow in the aorta of the dog rises abruptly during systole, its maximum coinciding with the peak of the pressure curve, and then declines, gradually at first and later suddenly. It reaches zero at the commencement of the protodiastolic period (p 209). Back flow occurs during the latter period. After this, forward flow is resumed and reaches a second maximum but not as great as the first, in the middle of diastole. Velocity of flow then declines gradually to the end of diastole.

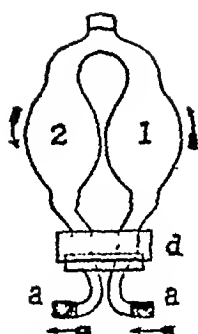


FIG. 17.1 Ludwig's stromuhr 1 and 2 glass flasks d , metal turn table, a , a section of artery. Arrows indicate direction of blood flow

it entails little or no disturbance of the circulation (fig. 17.2)

The velocity in the larger arteries of the dog during rest is from 0.1 to 0.2 meter per second, in the capillaries about 0.5 mm per second and in the medium-sized and large veins from 0.06 to 0.2 meter per second

There is no reliable method for determining the mean lineal velocity of the blood in man. Attempts are sometimes made to arrive indirectly at

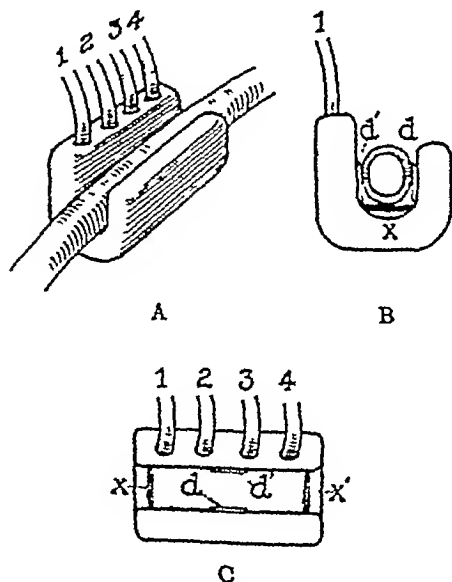


FIG. 17.2 Thermo-stromuhr *A* shows its application on a blood vessel, *B* shows a cross section of the blood vessel as it lies between the electrodes (platinum plates) d and d' , and on a thermojunction x *C* shows the relative positions of plates and thermojunctions x and x' are connected to a Zeiss loop galvanometer by leads 1 and 4 d and d' are connected to leads 2 and 3 (after Herrick, Esser and Baldes)

a rough approximation by the use of certain data obtained in other ways, e.g., output of the heart (ch. 26) and the cross section of the aorta in the cadaver (p. 143). The velocity in the latter vessel is from 0.25 to 0.30 meter per second during bodily rest. Estimations of the flow in the arteries of the arm may be made from determinations of the volume flow through the hand by the methods described below, and the diameter of the brachial,

CIRCULATION TIME

By this is meant the time which a particle of blood takes to make the complete round of both the systemic and pulmonary systems—*total circulation time*—or of the pulmonary circuit alone—the *pulmonary circulation time*. That is, if the voyage of a red cell could be timed from the moment it passed a given point, say in one jugular vein, to the moment when it arrived at the point from which it started or at a corresponding point in the opposite jugular the time consumed would be the total circulation time. The red cell must have traversed in succession the jugular vein, the right heart chambers, the pulmonary vessels, the left heart chambers, the arteries, and then the veins of the head and neck. The time taken in passing from the jugular to the carotid is the pulmonary circulation time,

Measurement of the circulation time This has been accomplished in *man* in various ways, e.g., by the injection of potassium ferrocyanide into the vein and having the blood from the other vein or artery fall upon a revolving drum covered with paper soaked in ferric chloride solution. The appearance of the Prussian blue reaction indicates the completion of the particular circuit which is being timed. *Methylene Blue* may be injected and the time noted when discoloration of the opposite vessel occurs. Stewart who employed the latter method also devised an *electrical method* in which a 2 per cent solution of sodium chloride is run into the vein. Two platinum electrodes separated from one another by a short distance are laid in contact with a vessel (jugular or carotid) of the opposite side. The electrodes are connected up as a resistance in one arm of a Wheatstone bridge. This is balanced with the resistance in the other arm. As the salt solution of high electrical conductivity reaches the section of vessel beneath the electrodes the electrical balance is upset by the sudden fall in resistance which results. This is recorded by the swing of a galvanometer or the ringing of a bell.

In man the circulation times have been measured by the injection of the dye *fluorescein* into

the arm vein of one side and timing its arrival at a corresponding vein of the opposite arm. This method, as originally introduced, required the withdrawal of blood at short intervals of time from the opposite vein and determining the time of arrival of the dye. The method has since been simplified by carrying out the experiment in a dark room and detecting its fluorescence in the minute vessels of the skin or mucosa membrane by means of a beam of ultraviolet light. The circulation time in man can also be measured by the intravenous injection of histamine (0.001 mg histamine phosphate per kilogram of body weight) and noting the time at which the flushing of the face, due to capillary dilatation, occurs, or by injecting a sweet or a bitter substance and having the subject signal the instant that he experiences the taste.*

Blumgart and Yens have devised a method in which an active deposit of radium is injected into the antecubital vein of one side and its arrival detected by means of an instrument sensitive to radioactivity, at the right auricle and at the antecubital artery of the opposite side. The detecting instrument must, of course, be shielded by means of lead sheets from the action of the radioactive substance except at these two points. The arm to arm time, less the arm to heart time, is taken as the "crude pulmonary circulation time". Since the time consumed in traversing the aorta, subclavian artery and the artery of the opposite arm is very brief (a fraction of a second) the measurement obtained in this way must approximate very closely the true pulmonary circulation time. A more convenient method for determining the pulmonary circulation time has recently been devised by Robb and Weiss based upon the fact that cyanide stimulates the respirations through its action upon the carotid sinus, 0.11 mg of sodium cyanide per kilogram of body weight is injected into the arm vein and the time recorded by means of a pneumograph and stop watch.

* A number of other substances have been employed for measuring the circulation time in man and animals, e.g., calcium gluconate (arm to tongue, average 12.5 minutes), which gives a hot sensation to back of tongue and throat, sodium dehydrocholate or decholin (arm to tongue), saccharin (arm to tongue), volatile substances, e.g., ether and perfumes (from arm vein to pulmonary capillaries), acetylcholine (leg vein to right auricle in animals, average 6.7 seconds), isotopic sodium Na^{24} , its arrival at any point in the circulation being detected by a Geiger-Mueller counter, and papaverin (arm to carotid sinus), which causes a deep inspiration upon reaching the carotid sinus receptors (p. 405).

NORMAL VALUES

- 1 Total circulation time Between veins of the two arms (fluorescein method) average 21 seconds (12 to 26 seconds) Arm to face (histamine method), average 24 seconds
- 2 Arm vein to corresponding artery of the opposite side (radio-active method), average 18 seconds (14 to 24 seconds)
- 3 Arm to heart (radio-active method), average 6.6 seconds (2 to 14 seconds)
- 4 Pulmonary circulation time (sodium cyanide method), average 10.6 seconds (7 to 14 seconds)
- 5 Arm to tongue (saccharin method) average 10.7 seconds (8 to 16 seconds)
- 6 Arm to face (histamine method) average 19 seconds (13 to 25 seconds)
- 7 Arm to lungs (ether method) average 6 seconds (3 to 8 seconds)

The circulation time is a measure of the shortest time which any particle of blood takes in passing from one point in the circulation to another. It is evident that if all the blood channels between the two points are not approximately uniform in length and diameter only a part of the blood will travel from point to point at the rate indicated by the measurement. It is believed, however, that the flow of blood through the lungs occurs at an approximately uniform rate in different vessels and that the pulmonary circulation time may be taken as an index of the mean velocity of the pulmonary blood flow. It therefore bears a relation to the total volume of blood traversing the lungs in a given time, i.e., to the output of the heart (p. 263), being short with large outputs and long when the output is small.

Knowing the pulmonary circulation time and the output of the heart per minute the volume of blood contained in the lungs at any moment may be calculated from the following equation—

$$C = Q (60/T)$$

where C = the cardiac output per minute, Q, the quantity of blood in the lungs and T the pulmonary circulation time. Q averages about 8 per cent of the total blood volume.

A close relationship exists between the size of the animal and the circulation times, these being shorter in a small than in a large animal. This is to be expected since the distances are shorter. The pulse rate also bears a relation to the animal's size, the smaller the animal the higher the rate. As a consequence, the product of the pulse rate and the pulmonary circulation time is almost the same in different species.

The circulation times are markedly reduced (velocity of blood increased) during muscular

exercise or by the injection of adrenaline. They are altered in the following pathological conditions

CONDITIONS ASSOCIATED WITH A REDUCED CIRCULATION TIME—

Hyperthyroidism The decrease is closely related to the height of the metabolic rate.

Anemia (see p 420) The increase in blood velocity is in proportion to the reduction in oxygen-carrying capacity of the blood—

CONDITIONS ASSOCIATED WITH A LENGTHENED CIRCULATION TIME

Hypertension, normal or slightly lengthened

Myxedema

Polycythemia vera

Cardiac failure The engorgement of the systemic and pulmonary vessels results in a great increase in the quantity of blood contained in the lung and peripheral vessels. The speed of the blood through these vascular areas is reduced though the total volume of blood flowing through the distended vessels in a unit of time is not necessarily diminished (see below). The improvement in cardiac action following the administration of digitalis is accompanied by shortening of the circulation time. This drug causes no effect upon blood velocity in normal persons—

Postoperative Smith and Allen found that starting about the 5th day after operation the foot to carotid circulation time increased gradually to the 10th day when it was 50 per cent greater than the preoperative value. These observations are pertinent to the question of postoperative thrombosis (see p 123)

THE VOLUME OF BLOOD FLOWING THROUGH AN ORGAN OF CIRCUMSCRIBED REGION OF THE BODY, THE MASS MOVEMENT OF THE BLOOD

The total volume of blood flowing through an organ in a given time should not be confused with the velocity of the blood stream in the individual vessels, i.e., the speed of the blood corpuscles which, as mentioned above, is referred to as the lineal velocity. The two do not alter to the same degree, or necessarily, indeed in the same direction. If the pressure head remains unchanged then the volume of liquid flowing in a unit of time along a narrow tube comparable in size to the small and medium-sized arteries is proportional to the fourth power of the radius of the tube. The lineal velocity, on the other hand, is proportional only to the sectional area of the tube (p 140). When

the arterioles of an organ dilate, though the peripheral resistance is reduced locally, compensatory vasoconstriction in other vascular areas occurs (*Löwen reflex*, p 288) and a fall in general blood pressure does not ordinarily result. That is, the pressure head in the vessels feeding the organ remains at, or may even rise above, its original level. A greater volume of blood therefore flows through the organ and the velocity of the blood in its larger vessels increases. The capillary blood pressure increases for, as a result of dilatation of the arterioles, less energy is expended in overcoming the frictional resistance in these vessels. If the capillaries do not alter in caliber, or do so to a minor extent when the arterioles dilate, the velocity of the blood flowing through the capillaries is also increased. If, on the other hand, the total cross area of the vascular bed of the organ is enlarged as a result of capillary dilatation an opposite effect upon the lineal velocity of the blood in these vessels is produced which tends to counteract the effect upon velocity caused by the arteriolar dilatation (velocity being inversely proportional to the cross area of the vascular bed, p 140). It thus comes about that with a constant arterial pressure the speed of the blood in the capillaries of a part may be reduced though the total volume of blood passing through the part remains unaltered or is even increased. In the case of the lungs, for example, whose vessels of course offer the only route for the blood from the right to the left side of the heart, dilatation or constriction of the arterioles and capillaries unaccompanied by any change in the output of the right ventricle will not alter the volume of the pulmonary blood flow. The velocity of the blood through the small vessels will, however, be reduced or increased respectively with the expansion or constriction of the pulmonary vascular bed. Conversely, owing to the occurrence of compensatory changes in the capacity of the vascular bed, a reduction or increase in the cardiac output (and so of the quantity of blood passing through the lungs or through the peripheral parts of the body) does not necessarily alter the velocity of the blood through the individual vessels of the pulmonary or systemic circuits.

Approximate values for the volume flow per 100 grams of tissue per minute for various organs under ordinary resting conditions of animals and man are given below

Thyroid
Kidney

560 cc.
150 cc.

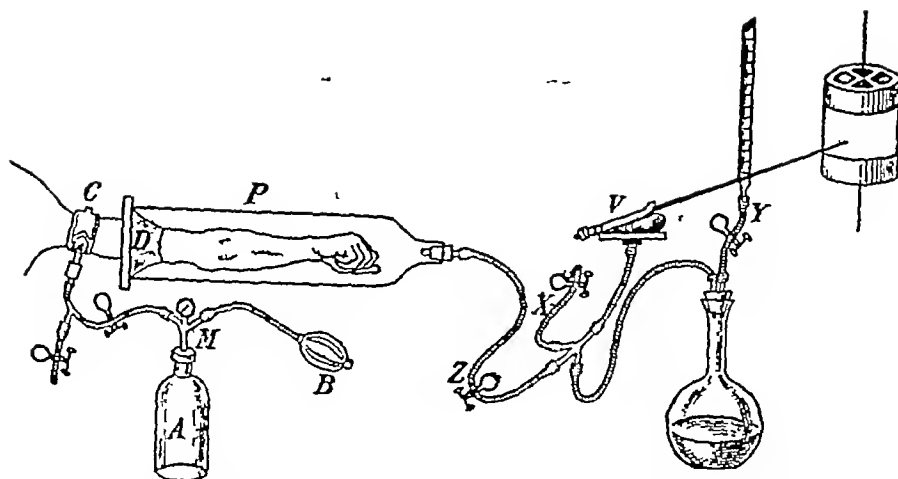


FIG 17 3 Diagram showing Hewlett and Van Zwaluwenburg's method for estimating the rate of blood flow in the arm (after Hewlett and Van Zwaluwenburg)

Liver	250 cc	{arterial, 50 cc portal, 200 cc
Brain (man)	200 cc	
Intestines	70 cc	
Spleen	40 cc	
Stomach	25 cc	
Human forearm (covered)	3 cc	
Skeletal muscle	3-5 cc	

Measurement of the volume flow

In *animals* one of three methods may be employed for the measurement of the volume flow through an organ

(1) The simplest method is to collect and measure the blood issuing from the vein or veins of the part. This method can be employed only over relatively short periods and with small well circumscribed organs such as the kidney or salivary gland.

(2) If there is a single artery or vein supplying or draining the organ the stromuhr may be inserted into either vessel and the quantity of blood measured which enters or leaves the region, or the thermostromuhr method may be employed (p 175).

(3) The plethysmographic method of Brodie. This is based upon the principle that if the venous return be occluded any change in volume of the part which results during the period of occlusion must represent the amount of blood which has entered the part during that period. The time during which the vein is compressed and consequently the duration of the observation must obviously be brief, for interference with the venous flow will automatically slow the blood stream and give a fallacious result. The organ, e.g., the kidney, with its blood vessels intact, is placed in an air-tight chamber (plethysmograph, p 294). At the point where the vessels enter and leave the chamber a soft material, e.g., sponge or tow packing smeared with vaseline, is used to form a hermetic seal, but does not compress the vessels. A tube leads from the interior of the chamber

to a tambour and recording apparatus. The excursions of the latter are calibrated to represent cubic centimeters of blood. To estimate the blood flow the vein is suddenly clamped and the increase in volume of the organ recorded over a short period. The method has been claimed by Brodie to give results comparable in accuracy with those obtained by means of the stromuhr. Brodie's method has been adapted by Hewlett and Van Zwaluwenburg to the estimation of the volume flow in the human hand and forearm. A narrow cuff encircling the upper arm is employed to compress the veins (fig 17 3). A plethysmograph for the finger on the same principle has been designed by Burton (fig 17 4).

The calorimetric method. This was devised by G. N. Stewart for the measurement of the volume flow through the human hand. Briefly, the method is based upon two assumptions: (1) that the hand is a perfect radiator, and when immersed in a small body of water the blood coursing through its vessels gives up heat, so that its temperature comes to equal the average temperature of the surrounding water during the course of the observation,³ (2) that the blood is the only source of heat, the heat generated by the muscular tissue of the hand being negligible.

The amount of heat (gram calories) given out by the hand to the water is obtained by multiplying the quantity of water in grams by the number of degrees centigrade through which its temperature has been raised. The blood flow in grams during the whole period of the observations is obtained from the following formula—

$$Q = \frac{H}{T - T_1} \times \frac{1}{S}$$

where Q equals the quantity of blood (grams), H the

³ That the temperature of the venous blood may be taken as that of the average temperature of the water in the calorimeter has been questioned by Sheard.

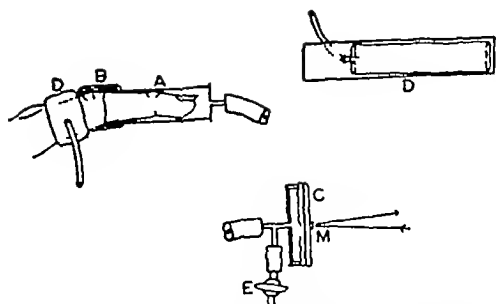


FIG 174 Finger plethysmograph (After Burton) *Up per figure*, A, metal cylinder into which the finger is inserted and sealed by means of a gauze binding (B) on zinc oxide paste, D, small cuff which when inflated occludes the venous outflow *Lower figure*, optical recording system C, capsule covered by a membrane of the thinnest rubber, M, mirror chip, E, by pass tap for re turning volume to the base line

beat (gram calories) given out to the water, T and T¹ the temperatures of the arterial and venous bloods respectively and S the specific heat of blood (0.8). The difference between T and T¹ represents the drop in temperature of blood in passing from the arterial to the venous side

A pair of specially designed water calorimeters of about 3000 cc capacity are employed for the observations.

The blood flow through the hand is thought to be indicative in a general way of the flow through the peripheral parts of the body

THE EFFECTS OF ACCELERATION UPON THE CIRCULATION

A constant speed, however great, has in itself no effect upon the circulation. As pointed out by Armstrong, we are travelling through space quite unaware of a speed of over 18 miles per second caused by the motion of the earth. But acceleration, that is, a change in velocity either in the line of our motion (linear acceleration and deceleration), or the movement of the body in a circular or curved course (centripetal or central acceleration), may cause profound effects upon the body as a result of the inertia of the blood and viscera. A third type of accelerated movement, known as angular acceleration, takes place when the body rotates about its own axis, as when a plane rolls or spins. The rate of angular acceleration is rarely great enough to cause any serious physiological disturbance, but this form of acceleration causes effects on equilibrium and orientation. The development of the modern airplane and maneuvers of military flying have brought prominently to the fore in recent years the importance and hazards of

acceleration in relation to the circulation. According to Newton's Law, $F = MA$ where F = force, M the mass and A the linear acceleration. Therefore, $A = F/M$

The effects of horizontal linear acceleration upon the body of a pilot are seen in catapulted take-offs, or in picking up personnel from the ground or in gliders by high speed planes. Deceleration, i.e., a sudden reduction in velocity, occurs in crash landings, parachute landings, or from parachute opening (especially at very high altitudes), etc. Vertical acceleration, as when the plane falls or pitches, owing to changes in air density, causes motion sickness (p 574)

Centripetal acceleration is the type which in military flying causes serious effects upon the circulation. It is defined as the acceleration of a body to ward the center of a circle in the circumference of which it is moving at uniform velocity. The central force producing the acceleration (centripetal force) is resisted by an equal and opposite, radially acting force, generally called centrifugal force. This may be expressed, thus —

$$G = MV^2/r$$

in which M is the mass, V the velocity and r the radius of the circular movement. At a linear or centripetal acceleration equal to that of gravity, namely, 32.2 feet per second, per second the force exerted is equal to that of gravity. Gravitational force, like centrifugal force, is proportional to mass and is exhibited as weight. The force due to acceleration is, therefore, conveniently expressed in units of gravitational force and designated by the letter G . Thus an acceleration amounting to $2G$ indicates a force double that of gravity, i.e., a body acted upon by such a force would be doubled in weight. The centrifugal force is proportional to the square of the velocity. It can be calculated in G units from the equation $G = V^2/32.2 r$, where V is the velocity in feet, r the radius in feet of the curved course which the moving body takes, and G units of gravitational force per pound of mass.¹⁰ Thus, the weight of a pilot's body subjected to a

¹⁰ From the following equation the force developed during linear acceleration or deceleration can be calculated

$$G = \frac{V_x^2 - V_y^2}{32.2 \times 2 \times S}$$

where V_y is the initial velocity, V_x the final velocity and S the distance travelled during the period of acceleration

force of 2G is doubled, is tripled at 3G, and quadrupled at 4G, and so on

Acceleration which produces a force acting upon the airman in the long axis of the body from head to seat is called positive (+G), that acting from seat to head is called negative (-G) A pilot pulling out of a power dive, that is, changing direction at high velocity to a horizontal and upward direction in a banking maneuver has his head directed inward toward the center of the circular movement and is therefore subjected to a positive centripetal acceleration If the force amounts to from 5 to 6 + G or more, and lasts for longer than 3 or 4 seconds, the phenomenon now generally referred to as "blackout" results, for his blood being acted upon by a force 5 or 6 times that of gravity "falls" or is "thrown" into the lower part of the body (the large vessels, it will be recalled, run in the general direction of the long axis of the body), the weight and consequently the hydrostatic pressure of the blood is increased (see fig 17 5) In animal experiments it has been shown that as a

result of the very high capillary pressure a marked increase in filtration of fluid from the blood into the extravascular spaces with hemoconcentration occurs In man the skin of the lower parts of the legs may show numerous blood extravasations (petechiae) The movement of blood toward the feet reduces the venous return of blood to the heart, and, as a consequence, the pressure of blood in the cerebral and retinal vessels falls Vision is temporarily lost and the pilot may become unconscious The abdominal viscera are forced downwards and drawing upon the diaphragm may embarrass respiration X-ray photography shows elongation of the heart and a reduction in cardiac volume

"Negative" acceleration, as when a turn is made at high velocity with the plane in the inverted position, the pilot's head being directed outward, causes opposite effects upon the circulation and is likely to produce more serious injury, but fortunately, unlike positive acceleration, does not come into play in any necessary maneuver "Negative" acceleration also occurs in spins re-

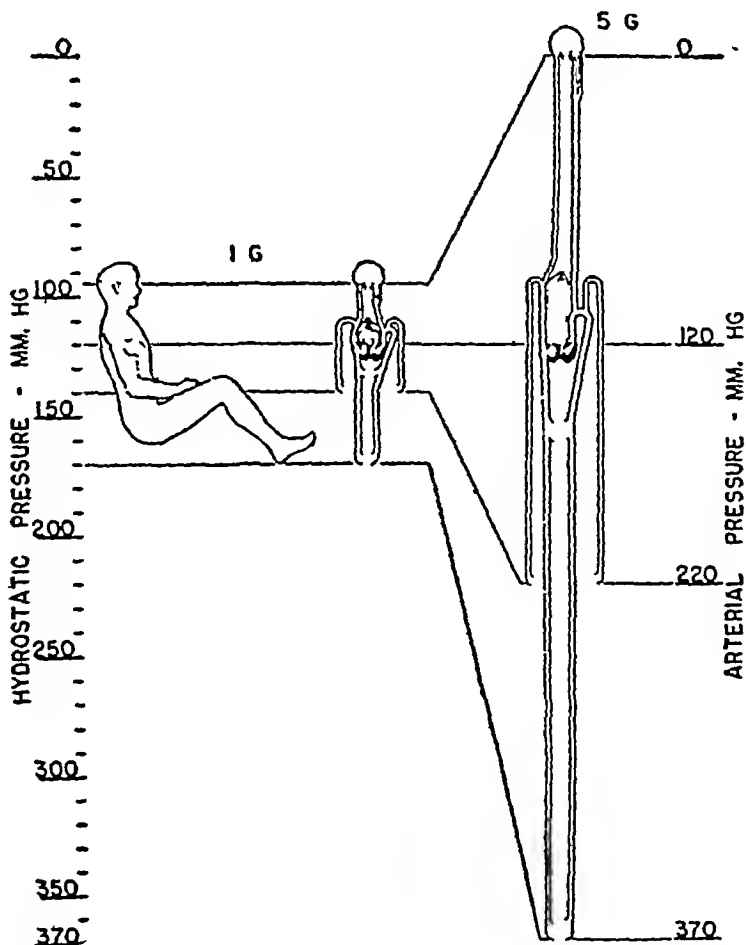


FIG 17.5 Diagrammatic representation of the hydrostatic pressures in the vascular system of a man in the sitting position at 1 G and at 5 G (Wood and associates)

sulting from some structural damage to the plane, when the tremendous force created may, through the increased weight of their bodies, prevent the pilot and other occupants of the plane from moving. In a turn with the head directed outward, the blood is driven toward the head resulting in effects sometimes called "red-out." The vessels of the head and neck become engorged, there may be small cutaneous hemorrhages, severe throbbing pain in the head is experienced and the eyes feel as though they were being extruded from their sockets. The abdominal viscera are pushed upwards against the diaphragm. The venous return to the heart is increased and the blood pressure in the cerebral vessels is raised considerably (as much as 65 mm Hg). There may be mental confusion for a time. Cerebral hemorrhage may result, but since the hydrostatic pressure of the cerebrospinal fluid is also increased, it acts as a buffer which, to a large extent, protects the delicate walls of the small cerebral veins and capillaries, which are the most likely to rupture.

Large centrifuge machines have been devised for subjecting the body to positive acceleration and studying the effects produced. Instruments record blood pressure, cardiac action and blood content of the drum membrane automatically. Thus the tolerance of a pilot to +G can be measured (see fig 176).

In order to prevent the physiological effects of positive acceleration, several types of double-walled suits have been designed which, by containing water or air under pressure, oppose a force to resist a rise in the hydrostatic pressure of the blood. The first suit of this kind to be employed in actual air combat was invented by Franks of the Royal Canadian Air Force. It envelops the abdomen and lower limbs, water is introduced into the space between its walls. During the development of centrifugal force a hydrostatic pressure, automatically graded to that exerted upon the blood, is applied through the tissues to the vascular walls. Thus, distension of the vascular bed of the lower

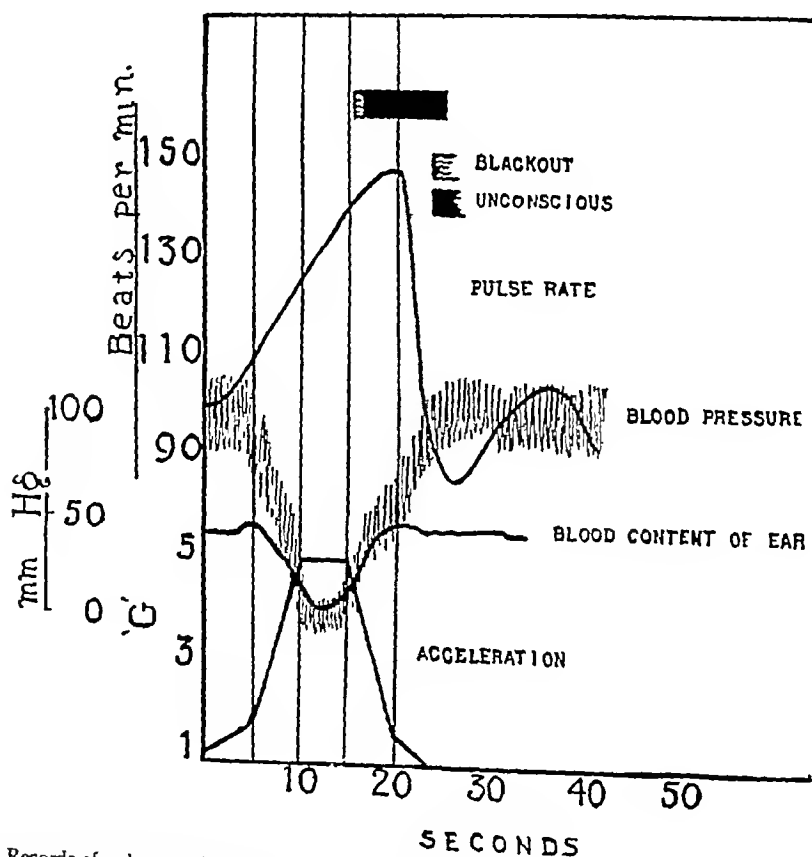


FIG 176 Records of pulse rate, blood pressure and blood content of the ear of a normal subject exposed in the human centrifuge for 5 seconds to 5 G (Drawn in part from graphs kindly furnished by Dr W. R. Franks.)

part of the body and the accumulation of blood is prevented

The effects of centripetal acceleration may also be minimized by the pilot assuming a crouching position with the trunk bent well forward and the thighs strongly flexed at the hips. The direction of

the centrifugal force is then across the great vessels of the trunk rather than in line with them, and runs from knee to hip more or less parallel to the vessels of the thigh, the movement of the blood along the veins of the thigh toward the heart is thus actually aided

CHAPTER 18

THE ARTERIAL PULSE

THE NATURE OF THE PULSE WAVE

The pulse is the pressure change created by the ejection of blood from the heart into the already full aorta and propagated as a wave through the blood column and arterial wall to the periphery. If the walls of the system were absolutely rigid, since liquids are incompressible, an impact delivered at one end would cause a pressure change to be transmitted instantaneously as through a steel rod to the furthestmost parts. In the case of elastic tubing, such as composes the vascular system, the pressure change is accompanied by an expansion of the tube's wall. It should be emphasized that the extra blood which is thrown into the aorta, and the pulse wave which is the direct result of the former, do not pass along the vessel in company. If the ejected blood were made distinguishable in some way and its speed compared with that of the pulse wave it would be found that the latter travelled at a rate from 10 to 15 times faster than the former. The speed of the blood depends upon factors which have been considered elsewhere, e.g., pressure gradient, cross sectional area of the vascular bed, etc., while the velocity of the pulse wave is determined almost entirely by the resilience of the arterial wall, the speed of the blood itself exerts a very minor effect.

The independence of the velocity of a liquid and of a wave in that liquid will be evident if one considers what occurs when, say, a bucket of water is thrown into a running stream. Waves are then set up which travel through the water at a rate which has little relation to the velocity of the stream itself. They run up the stream against the current as well as down. Though the pulse wave in its passage gives some forward movement to the blood, i.e., increases its velocity at the moment of its passage just as a wave upon a river slightly speeds the flow of water, as is shown by the impetus which it gives to a floating cork, yet the chief phenomenon is a propagation of a change of form and pressure. It is a molecular movement rather than a translation of fluid en masse. The particles of the fluid change their positions relative to one another, but to a comparatively small extent in relation to the arterial wall—a pulse wave travels through a perfectly stagnant blood column, as in a ligated vessel. This distinction may be illustrated by the diagram in figure 18 1.

MEASUREMENT OF THE SPEED AND LENGTH OF THE PULSE WAVE

The velocity of the pulse wave in the brachial artery is normally from 5 to 8 meters per second. By means of a pair of tambours, writing levers and a time marker (p. 226) the difference in the times of arrival of the pulse wave at a near and at a far point of the vascular system can be determined. The time difference divided into the distance travelled between the two points gives the velocity of propagation of the wave. Thus—

$$\frac{\text{Distance in mm}}{\text{Time in seconds}} = \text{velocity in mm per second}$$

Since the rate of propagation of the pulse alters with the state of the arterial wall, increased rigidity causing more rapid transmission and vice versa, the velocity is increased in old age or when the elasticity of the vessels is reduced by disease, e.g., arteriosclerosis. Bramwell, Hill and McSwiney found a speed of 5.2 meters per second at the age of 5, and one of 8.6 meters per second at 84 years. The velocity of the pulse in early adult life is about 6 meters per second. In hypertension (pathological or resulting from muscular exercise) the walls of the vessels are more strongly stretched and thus are nearer their elastic limit, the pulse wave is therefore transmitted at a higher speed. Vaso dilatation or low pressure from any cause, on the other hand, reduces the velocity of the pulse wave. The femoral artery is less distensible than the brachial and the velocity of the pulse is greater in the former vessel.

The length of the pulse wave. The length of a pulse wave tracing gives no indication of the length of the pulse wave itself. The tracing is inscribed by the rise and fall of the lever as the wave passes the point in the artery (fig. 18 2) and its length depends upon the rate at which the writing surface moves. When the rate is fast the curve is lengthened and vice versa. The pulse wave itself is in reality a very long swell or billow. The ejection of blood from the ventricle does not cause simply a small local dilatation but an extensive expansion of the arterial channels.¹ The wave measures from

¹ It is also to be remembered that the amplitude of the pulse wave is no criterion of the actual height of the blood pressure, for it may be low with a high blood pressure or vice versa. The pulse tracing or *sphygmogram* can give no absolute or quantitative information in respect to any phase of the arterial blood pressure. It is merely a record of the movements of the vessel walls brought about by pressure variations. The movements of the wall of a peripheral artery are

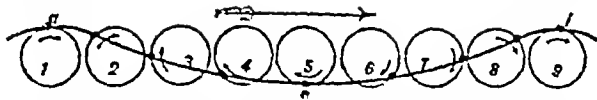


FIG 181 Motion in a water wave. Nine equidistant water particles are represented by dots. The particles move in vertical circles with fixed centers, all with the same uniform velocity. The position of each particle in its orbit is called the *phase* of its motion. In the diagram each particle differs in phase from the next by one eighth of a full revolution. At the top of the circle each particle moves in the direction of travel of the wave (as indicated by the large arrow), those towards the bottom move in the opposite direction. The particles upon the crests of the wave, *a* and *t*, are in the same phase and in a phase opposite to that of *c*, in the trough of the wave. After the wave has passed the particles come to rest and may show no change from their original positions. The length of the wave is the distance between two particles in the same phase (Modified from Kimball's *Textbook of Physics*)

3.50 to 5 meters and consequently the greater part has disappeared from the periphery before its end has left the aorta. The length of the wave may be determined from the following formula $L = VT$, where *L* equals length of wave, *V* the velocity of transmission and *T* the time the wave takes to pass any point. The manner of obtaining *V* has been described above. *T* may be obtained from the time markings on the pulse tracing or simply from the pulse rate. If, for example, 70 pulses pass a point in the artery per minute, each must require $\frac{1}{70}$ second for its passage.

ANALYSIS OF THE PULSE CURVE

The pulse wave may be made to describe a curve by laying a light lever upon an artery, e.g., the radial, and having it record its excursion upon a moving surface as the artery expands and recoils beneath it. The means of doing this is provided by an instrument known as a *sphygmograph*. One of the best known of these is Dudgeon's, shown in figure 183. The curve consists of an abrupt almost vertical upstroke, the *anacrotic limb*, and a more gradually sloping downstroke, the *catacrotic limb*. The rising limb is inscribed first and represents the front of the pulse wave, while the catacrotic limb is drawn during the fall of the lever, i.e., after the crest of the wave has passed. Though under abnormal conditions of the circulation (see below) the anacrotic limb may show one or more secondary waves, it is under ordinary circumstances smooth and uninterrupted. The descending limb, on the other hand, shows a well-marked negative

very slight, but in the sphygmograph tracing they are magnified several times. When the pulse is felt by the finger placed upon the radial artery it is the pressure change—the impact of the pulse wave—rather than the actual movement of the vessel wall which is detected.

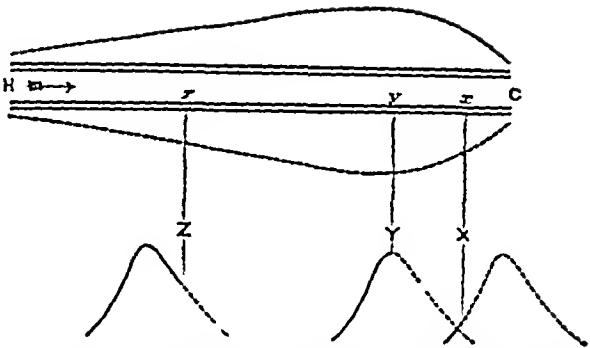


FIG 182 Diagram representing a pulse wave passing over an artery. Heavy double line, artery at rest, dotted line, artery expanded. Curves X, Y and Z are drawn during the passage of the wave (in the direction indicated by the arrow) by three levers placed upon the vessel at *x*, *y*, *z*. At *Z* the greater part of the wave has passed beneath the lever at *z*, as shown by the continuous line of the first curve, and has still to describe the smaller part represented by the broken line. In *Y* the lever at *y* is at the summit of the wave. At *X* the lever *x* has just commenced to rise and the greater part of the curve has still to be inscribed as indicated by the dotted line (after Foster)

wave, i.e., a depression, followed by a positive wave. These are known as the *dicrotic notch* and *dicrotic wave*, respectively. Under certain conditions the dicrotic wave may be so well-marked that it can be felt as a separate impulse by the finger placed upon the radial artery. Often less well-marked wavelets preceding the dicrotic notch (*predicrotic waves*) are present. Other waves (*postdicrotic*) may follow the dicrotic wave (fig 184).

The dicrotic wave and notch are produced in the following way. As ventricular systole comes to an end and the intra-ventricular pressure falls below the aortic, back eddies bring the aortic valves into apposition, at the same time the distended elastic arterial wall rebounds and pressing upon the blood forces it centrally as well as thrusting it forward toward the periphery. The swing of the

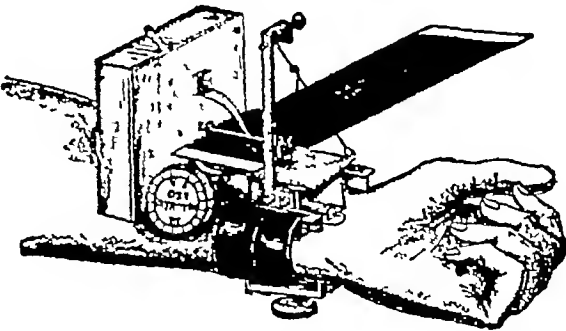


FIG 183 Dudgeon's sphygmograph, applied to the radial artery



FIG 18.4 The pulse wave. A. Primary wave, B. Dicrotic notch, C. Dicrotic wave

blood column toward the heart sets up a negative fluctuation which is propagated in the wake of the main or primary wave throughout the arterial tree and is represented by the dicrotic notch on the pulse tracing. The next instant the aortic valves which have been forced toward the ventricular cavity become taut, the movement of the blood column is abruptly checked and rebounds from their surfaces. This sets up a positive pressure change which appears in the pulse tracing as the dicrotic wave.

VARIATIONS IN THE FORM OF THE PULSE CURVE

The slope of the *ascending limb* is dependent upon several factors, (a) the duration of the ventricular discharge, (b) the output per beat, (c) the height of the diastolic pressure and (d) distensibility of the arterial walls. With a slow rate of ejection such as occurs in narrowing of the aortic ring (stenosis) the upstroke is gradual, the wave usually of low amplitude and a secondary fluctuation may appear (anacrotic wave)

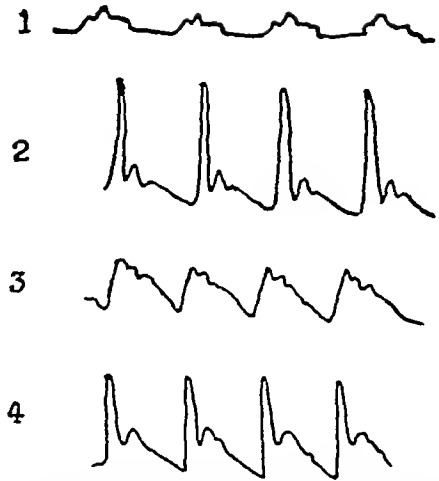


FIG 18.5 Abnormal pulse curves 1, aortic stenosis, 2, aortic regurgitation, 3, arterial hypertension, 4, arterial hypotension

which in some instances is actually higher than the primary wave (fig 18.5, 1). Secondary waves upon the anacrotic limb are also found in aortic aneurysm or when the aorta is narrowed as a result of a developmental defect or by pressure from without. These waves on the anacrotic limb are probably reflected from the periphery of the circulation, the long period of ejection permitting waves to travel centrally and meet the primary wave before it has reached its crest.

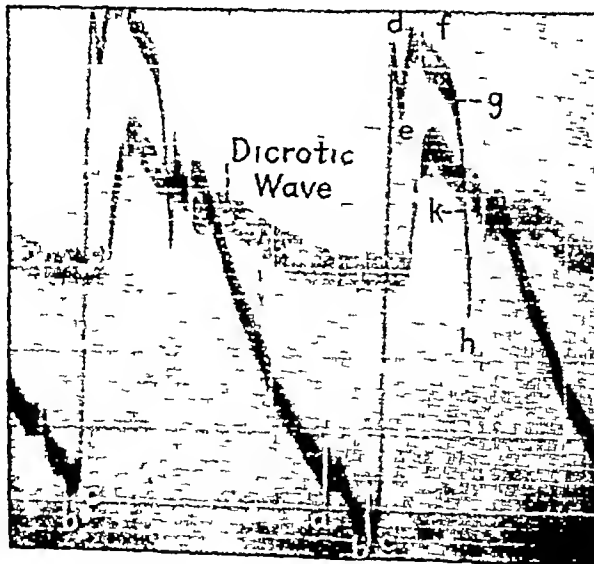


FIG 18.6 Tracing showing the difference between the subclavian (central) and radial (peripheral) pulse in man. The delay of the radial pulse is clearly shown (after Wiggers)

When the discharge is rapid, the output large, or the diastolic pressure low, the pulse wave tends to be of large amplitude and its upstroke steep (fig. 18-5, 1). The slope of the *descent limb* is abrupt when the diastolic pressure is low, as in aortic regurgitation or arteriovenous aneurysm (fig. 18-5, 2).

Any state which renders the arterial walls less tense will tend to increase the magnitude of the diastolic fluctuations and to produce other secondary waves (pre-diastolic and post-diastolic) which under ordinary circumstances do not appear. If, for example, the arterial system is underfilled, as a result of vasodilatation or a fall in systolic pressure from whatever cause exists, the diastolic wave becomes prominent, under such conditions as in typhoid fever, it may sometimes be felt with the finger as a distinct tap following the primary wave. The pulse is then said to be diastolic.

THE PULSE WAVE IN DIFFERENT REGIONS CENTRAL ARTERIAL PULSE

Pulse tracings recorded at different parts of the arterial tree show that the configuration of the wave changes as it travels toward the periphery. The rise and fall of the primary wave are more gradual and the amplitude is less in the smaller arteries. A pulse curve taken from a central artery, such as the subclavian, differs very strikingly from the typical tracing taken from a peripheral artery, such as the radial. Instead of the usual comparatively shallow diastolic notch and

rounded wave, a sharp depression (*incisura*) followed by a sharp spike is seen in the subclavian tracing. Such a record may be obtained from the human subclavian by means of a small conical cup pressed into the supraclavicular fossa and connected with an optical recording apparatus (fig. 18-6). This *central arterial pulse* shows two preliminary vibrations. These immediately precede the main upstroke, i.e., they occur before the opening of the semilunar valves, and result from pressure changes transmitted through these delicate structures. The first (a-b) is due to auricular systole, the second (b-c) to the tension developed at the beginning of ventricular systole (isometric period, p. 200). The ejection of blood from the ventricle causes a sharp rise in pressure which, setting the blood column into vibration, causes the fluctuations c-d-e. The pressure mounts to f and is sustained for a time, but then as systole comes to an end it drops precipitately to produce a deep depression—the *incisura* (g-h). This occurs during the closure of the semilunar valves and corresponds to the so-called protodiastolic period of ventricular diastole (p. 200). Several waves follow the *incisura* due to after vibrations of the valves. In their transmission through the arterial wall to peripheral points minor fluctuations of the central pulse become fused to produce smooth contours, while the primary oscillations, the *incisura*, and the after vibrations become damped down and appear eventually as the primary wave and diastolic fluctuations of the distal arterial tracings.

THE PHYSIOLOGY OF CARDIAC MUSCLE PERFUSION OF THE ISOLATED HEART

HISTOLOGY OF CARDIAC MUSCLE

Cardiac muscle fibers, like those of voluntary muscle, have transverse as well as longitudinal striations. But the investing sarcolemmal sheaths of the cardiac fibers, unlike those of skeletal muscle, are ill-defined and the muscle cells communicate with one another through branches of bridges of protoplasm (fig 191). So, the cardiac muscle, rather than being a collection of separate fibers, has the characters of a syncytium—a continuous multinucleated sheet over which an impulse may spread without interruption in any direction. The cross striations are less distinct than in skeletal muscle, the sarcoplasm is more abundant and the nuclei are embedded in the center of the cell. Transverse bands, known as *intercalated discs*, are disposed at irregular intervals in cardiac muscle. Their function is unknown. Cardiac muscle fibers have no power of regeneration, once destroyed they are replaced by fibrous tissue only.

THE PHYSIOLOGICAL PROPERTIES OF HEART MUSCLE

These may be considered under the following heads (1) *Excitability and contractility*, (2) *rhythmicity*, and (3) *conductivity*.

(1) *Excitability and contractility*

The ability of a tissue to respond to a stimulus is spoken of as *excitability* or *irritability*. In the case of muscle the response is a shortening of its fibers. Certain features of the contraction of cardiac muscle will now be discussed.

(a) **THE "ALL OR NONE" LAW** This states that the weakest stimulus that is capable of causing a contraction at all (minimal stimulus) will produce the maximal contraction. This fundamental fact, which was first demonstrated by Bowditch in 1871 in cardiac muscle, has since been found to be true also for nerve and skeletal muscle. Though a skeletal muscle responds to stimuli of graded strengths by contractions of graded amplitude and so apparently differs from cardiac muscle, it can be shown that a single skeletal fiber obeys the

"all or none" law. The apparent difference in behavior between cardiac and skeletal muscle is due to the fact that the former is a continuous protoplasmic sheet, and an impulse which causes contraction in one part spreads under ordinary circumstances and involves the whole. The individual fibers in skeletal muscle, on the other hand, are insulated from one another. The graded response of the entire skeletal muscle means simply that more fibers are excited by the stronger than by the weaker stimulus.

It must be pointed out that a minimal stimulus at one time may at another be subminimal should the excitability of the muscle be reduced. On the other hand, should the excitability be increased an ineffective stimulus may be rendered effective. Consequently, when it is said that the cardiac muscle follows the "all or none" law it is to be remembered that this applies only to the conditions existing at the time. The excitability and contractility of the muscle are variable and a stimulus which under one set of conditions would produce a weak contraction, would under more favorable conditions produce a much greater response. Length of the fiber, hydrogen ion concentration, state of nutrition, fatigue and the inorganic constitution of the fluid bathing the heart (p 192) are among the factors influencing its excitability and its contractile force.

(b) **TREPPE** This phenomenon was also first observed by Bowditch in cardiac muscle, though it is also shown by skeletal muscle. If a number of stimuli of the same intensity (maximal) be sent into the muscle after a resting period, the first few contractions of the series increase successively in amplitude. This ascent in the magnitude of the responses at the beginning of the series suggests the rising steps of a staircase (German *treppe*) and is supposed to be due to the greater contractility of the muscle at this time, resulting from the rise in temperature and the slight increase in H⁺ ion concentration of the muscle as a result of lactic acid production. This apparent contradiction of the "all or none" law is explained by the considerations in the foregoing paragraph.

(c) **THE REFRACTORY PERIOD** The refractory

period of skeletal muscle is very brief (0.01 second) and corresponds to the latent period, i.e., the period elapsing between the receipt of the stimulus and the commencement of the contraction. A stimulus applied to the muscle any time after it has commenced to contract causes a second contraction which is added to the first (*summation*). A rapid series of stimuli timed to fall in each instance just after the refractory period of the preceding contraction will produce a rapid series of shortenings of the muscle which fuse into an apparently maximal contraction (*tetanus*). The contraction is sustained as long as the stimulation is continued or until fatigue sets in. Heart muscle behaves differently. Its refractory period is relatively long and lasts throughout the contraction phase. *The muscle will not respond to a second stimulus, no matter how strong, so long as its fibers are still in the contracted state*¹. This is the *absolute refractory period*.

The time interval following a previous excitation during which a stimulus will not elicit an impulse of sufficient intensity to be conducted through the cardiac muscle is termed by Drury the "*effective refractory period*".

During relaxation the muscle regains its excitability gradually. Early in the relaxation phase the strength of stimulus required to produce a response is greater and the response to a given stimulus is less than later on. Not until complete relaxation has occurred does the excitability of the muscle return to its normal value. This period of time during which excitability is depressed but not abolished is called the *relative refractory period*.

The latent period of the response, that is, the time elapsing from the application of the stimulus to the contraction of the ventricle, is of the same length whether the latter is induced early or late in the relative refractory period (Woodworth, Fiddes). The long latent period when the stimulus was applied early after the previous beat, as found by Marey and shown in his tracing (fig. 19 2), was due, it now appears, to the inadvertent stimulation of the auricle by the escape of current while the ventricle was absolutely refractory. The impulse arising in the auricle by the time it has reached the ventricle finds it again responsive and an extrasystole results. The long latent period in these

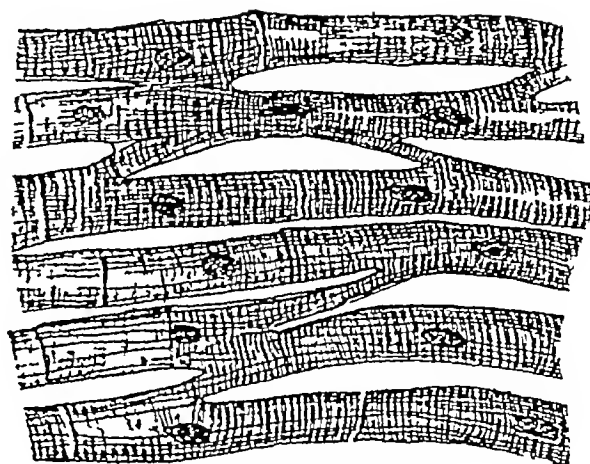


FIG 19 1 Cardiac muscle fibers

classical experiments therefore includes the latent period of the auricle as well as the time taken for the impulse to reach the ventricle.

It will be noted from a reference to figure 19 2 that a long pause follows the contraction caused

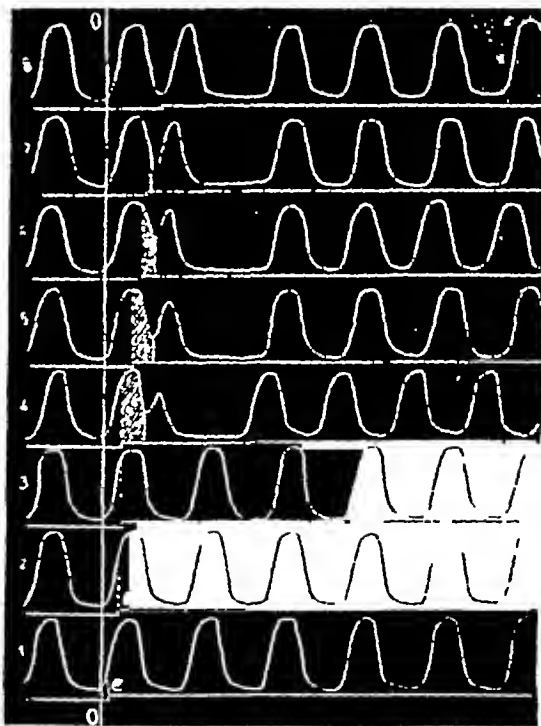


FIG 19 2 Myograms of frog's ventricle, showing effect of excitation by break induction shocks at various moments of the cardiac cycle. The line 0, 0 indicates in all tracings the commencement of the beats during which the shocks were sent in. It will be noted that in 1, 2 and 3 the heart is refractory to the stimulus. The signal (the break in the horizontal line) indicates the moment at which the stimulus was applied. The latent period (hatched area) does not alter as this figure shows. See text. The extrasystoles increase in height from 4 to 8, each being followed by a compensatory pause (after Marey).

¹ Wiggers finds that the refractory period of the heart does not last for the entire duration of its contraction but that it responds by a contraction early in diastole to a stimulus applied within the last 0.06 second of systole.

by the artificial stimulus. The artificially induced contraction is called an *extra-systole* or *premature contraction*. The long interval following the *extra systole* is termed the *compensatory pause* since its duration is such that when the next normal beat is resumed it occurs at precisely the same time, i.e., at the same point in the tracing, as it would have appeared had no premature contraction been provoked. The occurrence of the long pause is explained as follows. The normal impulses pass from the auricle to the ventricle in a perpetual stream and in orderly sequence. When the ventricular muscle is stimulated artificially during diastole, and an extra contraction induced, the normal impulse when it arrives from the auricle at the usual time finds the ventricle already in the contracted state and, in consequence, refractory. The impulse is therefore ineffective. Not until the arrival of the next normal impulse is the muscle in a condition to respond.² This accounts for the fact that the time elapsing between the normal beats preceding and following, respectively, the premature contraction is equal to the length of two normal cardiac cycles. In other words, the heart after the interruption in its rhythm again "gets into step."

The long refractory periods of the cardiac muscle serve to preserve the cardiac rhythm. The absolute refractory phase makes the summation of contractions and the production of tetanus impossible. The relative refractory period tends to discourage the occurrence of a second contraction before sufficient time has elapsed to allow the complete relaxation of the muscle from a preceding

² Sometimes when the premature contraction occurs early in diastole, being very weak and short, it is over before the next normal impulse arrives. Therefore a long pause does not appear. An extra contraction situated between two normal beats and not followed by a long pause is called an *interpolated extra systole*. The underlying processes responsible for the refractory periods, absolute and relative, are unknown. They have, however, been given a somewhat picturesque explanation based upon the assumption that the cardiac contraction results from the liberation of energy accumulated during the diastolic period. During systole this energy is "touched off" by the cardiac impulse and discharged. At the end of the contraction phase it has been completely dissipated, a second stimulus is therefore ineffective. In early diastole a relatively small amount of energy has accumulated and a stimulus applied at this time in consequence calls forth a very weak response. Subsequent responses correspond to the amount of energy built up between the previous systole and the application of the stimulus. This conception is useful for purposes of illustration but it gives little aid in gaining an insight into the fundamental processes concerned.

contraction. When a premature contraction does occur its refractory period serves to restore the normal rhythm.

The absolute refractory period may be altered by various agencies. It is shortened by a rise in temperature and by rapid heart action. It is prolonged by the action of certain drugs (p. 236). Vagal stimulation shortens the refractory period of the auricular muscle but has no effect upon that of the ventricular muscle (p. 242).

(2) *Rhythmicity* and (3) *conductivity*

These properties of cardiac muscle will be considered in subsequent pages.

Cardiac tonus

The subject of tonus in mammalian cardiac muscle has been beset with conflicting opinions. Rhythmical tonus changes were demonstrated by Fano in the auricle of the tortoise. This observation has been repeatedly confirmed, yet it should not be cited in support of the view that tonus is a property of the mammalian heart since the rhythmic changes were shown by Bottazzi to arise in a sheet of smooth muscle lying beneath the endocardium and not in the cardiac fibers proper. Tissue of this nature is absent from the warm-blooded heart and from the ventricle of cold-blooded animals. Tonus in the ventricles of the latter as well as in either chamber of the mammalian heart has yet to be demonstrated.

A great deal of confusion has arisen regarding the question of cardiac tone, through different interpretations having been given to the term itself. Tonus of muscle in the usual physiological sense means that state of partial and sustained contraction by which the muscle offers resistance to being stretched above or in addition to that which is offered by its purely physical properties. For instance, a skeletal muscle at rest and in connection with the central nervous system exhibits a slight but definite and persistent contraction. Less force is required to lengthen it when the nerves are cut and the tonic contraction disappears. The resistance which the muscle offers after this is dependent purely upon its inherent physical qualities, such as may be possessed by non-viable material, e.g., an elastic band.

With regard to the muscle of the heart the question is this: During diastole, does a certain degree of slight contraction—tonus—persist which offers some resistance to the inflowing blood and in consequence influences at this time the length to which the fiber is stretched? There is little definite evidence that tonus in this sense is possessed by the mammalian heart.

muscle, but variations in extensibility of cardiac muscle are admittedly more difficult of investigation than are those of skeletal muscle. In the intact heart the only means available for such a determination is the detection of a change in the volume of the heart when a single experimental condition which might be expected to affect tonus is altered while all other factors are kept at constant values. The capacity of the ventricular cavity can increase, of course, only by a lengthening of the cardiac fibers, therefore the volume of the ventricle during diastole (diastolic volume) will give during this period an index of fiber length. Variations in the latter can be taken to represent tonus changes only when other influences which will affect the diastolic volume are eliminated. For instance, increase or decrease in the venous return during the experiment would alter the fiber length, variations in the duration of diastole would also influence the diastolic volume, since more blood will flow into the ventricle during a long than during a short diastolic period. These factors must therefore be adequately controlled before justifiable conclusions concerning tonus changes can be drawn. Apparently the constancy of the experimental conditions has not always been assured and erroneous judgments upon cardiac tonus have been pronounced.

Evidence of tonus changes derived from the action of drugs upon the heart are inadmissible since abnormal conditions may be thereby induced.

It has been customary to use the term "tone" somewhat loosely, and to refer to a heart showing a large or small diastolic volume as possessing "low tone" or "high tone" respectively. But as just stated the difference in volume may be the result of other factors and need have nothing to do with tone.

PERFUSION OF THE ISOLATED HEART, NUTRITION AND OXYGEN REQUIREMENT OF THE CARDIAC MUSCLE

The frog's heart possesses no circulatory system, but merely obtains its oxygen and the necessary nutrient materials from the fluid which bathes it. In order to perfuse successfully the excised heart of higher animals certain special conditions must be fulfilled. When this has been done the mammalian heart (not excepting the heart of the human subject excised shortly after death) may continue to beat for several hours.*

The following requirements must be satisfied in carrying out the perfusion of the mammalian heart.

(1) Pressure

It is essential that the perfusion fluid be delivered under sufficient pressure in order to drive it

* Mann and his colleagues have succeeded in transplanting the heart from one dog to another and having it beat regularly for eight days.

through the coronary circulation. This may be accomplished most simply by means of a reservoir raised to a height equivalent to the normal aortic blood pressure (i.e., about 5 feet). Following the method of Langendorff a cannula is tied into the aorta (fig 19.3). The heart is allowed to hang from the cannula which is fixed to a support and connected by tubing to the supplying reservoir. The aortic valves are brought naturally into apposition by the pressure of the fluid, which then enters the coronary arteries. Little or no fluid enters the cavity of the left ventricle directly from the aorta, after completing the coronary circuit it escapes into the right auricle by the coronary sinus and anterior cardiac veins. A smaller amount enters the ventricles through the veins of Thebesius and other channels.

(2) Temperature

The perfusion fluid is maintained at body temperature by passing it through a glass or metal worm immersed in a water bath whose temperature can be regulated by a thermostat. Increase or decrease in temperature causes a corresponding change in heart rate. The rate of the frog's heart is doubled by a 10°C rise in temperature between 4° and 20°C. That is, the temperature coefficient

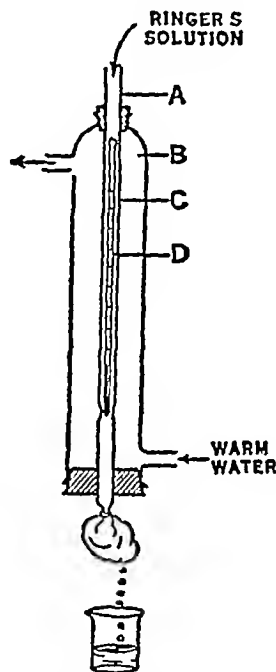


FIG 19.3 Showing a simple apparatus for the perfusion of the mammalian heart. Ringer's solution flows down the narrow tube under a pressure of about 120 mm of mercury. The lower end of this tube is tied into the aorta so that the fluid will flow into the coronary arteries. Surrounding the narrow tube is a glass water jacket through which warm water circulates.

is about 2 (fig 19 4) A corresponding relationship between heart rate and temperature holds for the mammalian heart within the temperature range from 26° and 40°C

(3) Oxygen supply

This is furnished by bubbling oxygen or compressed air through the perfusion fluid The oxygen consumption of the heart in a heart-lung preparation under conditions approximately those of bodily rest was found by Evans and Starling to average 3 24 cc. per gram of heart tissue per hour, but with the heart in situ, and under more physiological conditions, a figure of 5 3 cc. per gram per hour has been obtained for the left ventricle of the anesthetized dog The oxygen consumption of the isolated heart is probably much less than this During maximum work, on the other hand, the heart of the intact animal probably consumes oxygen at the rate of 30 cc. or more per gram per hour Experiments with the heart-lung preparation show that the oxygen usage is directly proportional to the work performed by the heart, which in turn runs parallel with the length of the muscle fiber (p 255) The oxygen consumption of the heart is greatly increased by the addition of adrenaline to the perfusion fluid The efficiency of the heart is also increased by adrenaline in physiological dosage

Cardiac muscle can run up but a small oxygen debt (p 725) and ceases to function usually

when the debt amounts to 0 06 cc per gram of tissue, that is, about a fifth of the debt which skeletal muscle can contract (0 30 cc per gram) The heart muscle is highly sensitive to changes in pH and has a much lower buffering power than has skeletal muscle As a consequence, when the lactic acid in heart muscle reaches a concentration of about 0 07 per cent or the pH falls to around 6 25, extra-systoles, heart block or other irregularities occur, followed by cessation of the beat Skeletal muscle, on the other hand, continues to contract until a lactic acid concentration of over 0 2 per cent is reached The mammalian heart is unable to do work for more than 5 or 10 minutes after its oxygen supply has been cut off

(4) Chemical constituents

Blood serum or defibrinated blood, or whole blood to which hirudin or heparin has been added, may be employed as the perfusion fluid An artificial fluid, however, is frequently used but it must imitate the plasma in so far as the chief inorganic salts of the latter are concerned

Such a solution is that first devised by Ringer, who drew attention to the importance of the three cations, Na⁺, K⁺ and Ca⁺⁺ in the same proportions as they exist in plasma, for the maintenance of the normal action of the heart

Ringer's solution contains the three elements in the form of the chlorides of sodium, potassium and calcium Several modifications of this original fluid have been made *Locke's* solution, for instance, contains in addition sodium bicarbonate and glucose *Tyrode's* solution is similar to the former but also contains a small percentage of magnesium chloride and of the acid and sodium phosphates The phosphates are designed to give the solution an optimum concentration in H⁺ ions Table 18 gives the percentage compositions of the

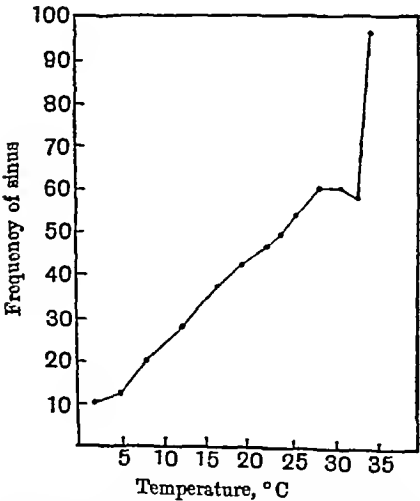


FIG 19 4 Showing the linear relation of temperature to heart rate over a range of about 30°C in South African frog (Taylor)

TABLE 18
Percentage composition of perfusion fluids

	RINGER'S (MODIFIED FOR MAMMALIA HEART)	LOCKE'S	TYRODE'S
NaCl	0 900	0 900	0 900
KCl	0 040	0 042	0 02
CaCl	0 025	0 024	0 02
MgCl			0 01
Glucose		0 1	0 1
NaHCO ₃	0 020	0 020	0 1
NaH ₂ PO ₄			0 005

three different fluids. The nitrogenous constituents of blood serum, e.g., proteins, urea, etc. apparently have little effect upon the beat.

Clarke found, however, that a frog's heart after perfusion for some hours passed into a "hypodynamic" state. Contraction and conduction became greatly impaired. This condition was attributed to the removal of lipid materials from the muscle cells. The addition of certain lipid materials partially restored the heart to its previous condition.

THE ACTIONS OF THE DIFFERENT CATIONS UPON THE HEART BEAT

Ringer observed that if the heart were perfused with 0.6 per cent sodium chloride solution a few beats were executed, but the heart then stopped in diastole. The addition of calcium restored the beat for a time but the heart again came to a standstill, this time in systole. The addition of potassium antagonized the calcium effect, the beat recommenced and was maintained.

It is now well known that calcium in excess, or

in normal concentration, but in the absence of potassium, lengthens systole at the expense of diastole. The heart finally stops in the fully contracted state—*calcium rigor*. Potassium acts in a reverse manner if in excess or unbalanced by calcium. More and more of the cardiac cycle is occupied by diastole and the heart ultimately comes to rest in the completely relaxed state—*potassium inhibition*. A solution containing calcium and potassium alone will not sustain the beat, sodium is essential. The manner in which sodium acts is not so clearly demonstrable as in the case of the other cations, but it is certain that the excitability and contractility of the heart muscle cannot be maintained in its absence (fig. 19.5).

It is apparent then that these three cations are absolutely necessary for the normal beat of the heart, the calcium increasing the contractility and prolonging systole. Potassium has the reverse effect, reducing contractility and favoring relaxation. The presence of these substances in

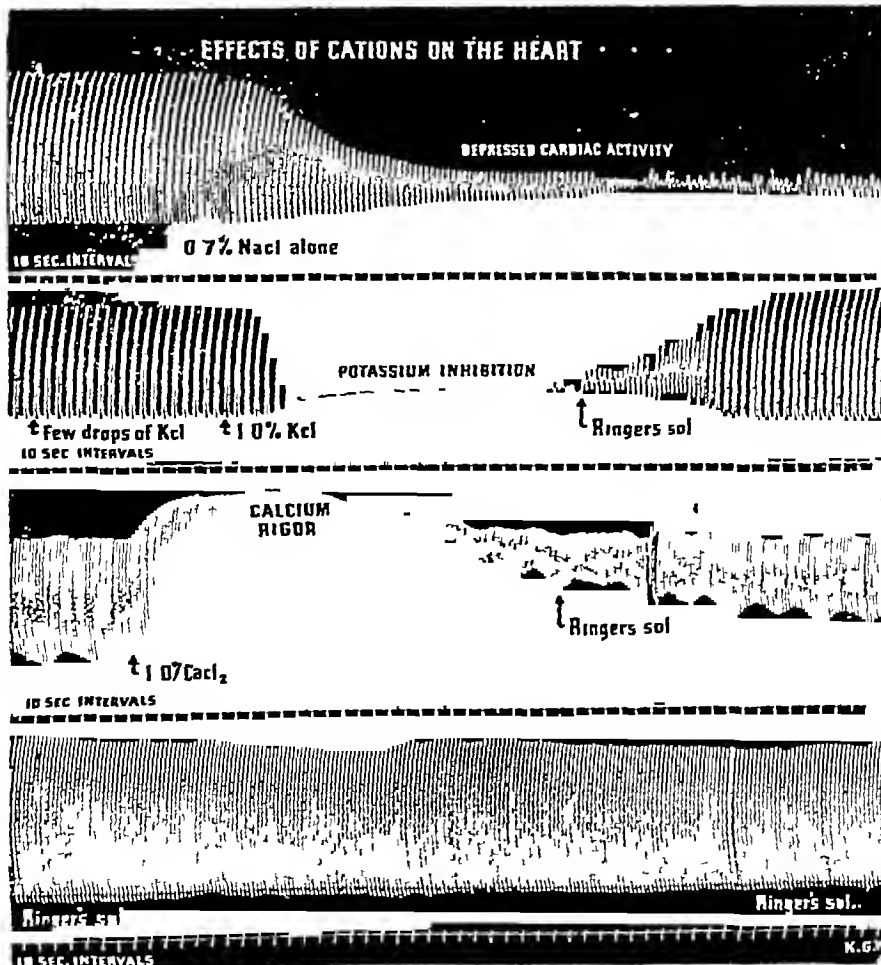


FIG. 19.5 Showing the effects of cations on cardiac behavior (Kindness of Dr. K. G. Watkin)

proper proportion ensures the rhythmicity of the contractions. It would appear, however, that potassium is of less importance than the other two, since the turtle's heart will beat for a time when this ion is absent. On the other hand, this may simply be due to the well known fact that the muscle cells have themselves a rich store of potassium. With regard to other bivalent cations, strontium, though less efficient, can replace calcium in the perfusion fluid, but barium is toxic, and magnesium is inert.

Keith has studied the effects of potassium salts upon the human heart. The characteristic change observed in the electrocardiogram, following the administration of a large dose of potassium (60 to 100 mg per kilogram of body weight), was a high peaked T wave. Even much smaller doses produced this effect, though to a less pronounced degree.

The effect of potassium in producing fibrillation in the mammalian heart is described on p 238.

Of the underlying physical or chemical changes through which these elements influence the heart beat little is definitely known. Calcium decreases and potassium increases the permeability of the cell membrane. According to some, the excitability and contractibility of the muscle fiber is dependent upon the relative concentrations of H ions on the two sides of the cell membrane. The Ca and K ions through altering the cell permeability may in this way affect the diffusion of H ions across the membrane and so vary their relative concentrations and as a result the magnitude of the differences in electrical potential on the two sides.

THE EFFECT UPON THE HEART BEAT OF CHANGES IN THE REACTION OF THE PERFUSION FLUID

Acids when added in *moderate* excess act like potassium in that they favor relaxation of the cardiac muscle. The heart finally comes to rest in diastole. Alkalis, on the other hand, serve to prolong systole and shorten diastole. In this way they act like calcium. Acids depress and alkalis increase conductivity through the auriculo-ventricular bundle (p 199). Andrus and Carter found that when the pH of the fluid perfusing the isolated mammalian heart was reduced to 7.0 complete heart block occurred (p 227).

The higher tension of carbon dioxide during muscular exercise probably exerts a beneficial effect at this time upon cardiac behavior. During exercise the blood flow through the muscles is

greatly augmented and a larger volume of blood is returned to the right side of the heart, the rise in carbon dioxide tension will favor more complete relaxation of the cardiac muscle for the accommodation of the greater load of venous blood.

THE METABOLISM OF CARDIAC MUSCLE

In as much as the myocardium derives the energy for its contraction from the oxidation of food material, it must receive continuously an adequate supply of oxygen. The heart beats for only a very brief period after its blood supply has been cut off. The heart muscle can go into debt for an amount of oxygen which is only about $\frac{1}{3}$ of the oxygen debt possible for skeletal muscle—around 0.06 gram of oxygen per gram of tissue. The metabolism of the myocardium rises, as is to be expected, with increased work, whether the extra work is entailed by a rise in arterial pressure, i.e., by increased resistance to the discharge of blood, or to the greater amount of blood which it receives (venous inflow) and ejects (cardiac output). In the case of the heart with its nervous control intact, the work is done in the former instance with higher efficiency. But in a heart deprived of nervous control, as in the heart-lung preparation (p 251), increased work due to raised arterial pressure is performed with less economy of oxygen usage—that is with less efficiency—than when an equal increase in work is the result of increasing the load of venous blood. Increase in heart rate itself increases the metabolism of the heart muscle, the heart with its nerves intact slows when the arterial pressure is raised (Marey's Law) but quickens when the venous inflow is increased (Bainbridge reflex). These alterations in heart rate, which do not occur in the heart whose nervous regulation has been abolished, explain in part but not entirely, the relatively high efficiency of the innervated heart in performing extra work due to raised arterial blood pressure. The effect of nervous control upon cardiac efficiency appears to be bound up in some way with the action of the cardiac nerves (vagus and sympathetic) themselves upon the metabolism of the heart muscle, the vagal effect being toward a reduction in oxygen consumption, whereas the sympathetic nerves and adrenaline increase cardiac metabolism quite apart from their effects upon the heart rate. The vagal effect may be so pronounced that in performing work against a raised arterial pressure the heart may actually use less oxygen than at a lower pressure.

THE FUEL OF THE CARDIAC MUSCLE

From experiments in the past it had been concluded that the heart muscle derived energy by the direct utilization of blood sugar, since glucose disappeared from blood perfusing a heart-lung preparation, or from fluid used to perfuse the isolated heart. Figures ranging from 0.8 to 5.3 mg per gram of heart muscle

per hour have been obtained by different observers for the quantity of glucose removed by the heart. It has been shown, however, that a rapid breakdown of glucose (glycolysis) with the production of lactic acid occurs in blood after its removal from the body, i.e., free from contact with any tissue. Evans and his associates found that lactic acid was produced, as a result of glycolysis, at the rate of about 14 mg per 100 cc of blood per hour. Glycolysis also occurs in an artificial perfusion fluid as a result of bacterial action, and it has been shown that if rigid precautions are taken against bacterial growth, relatively small quantities of glucose disappear from fluid used to perfuse the excised heart.

The work of McGinty and of Lovatt Evans and their associates indicates that the heart muscle utilizes glucose to a minor extent. On the other hand, the heart removes relatively large amounts of lactic acid from the blood. McGinty and Miller analyzed the ingoing (arterial) and outgoing (venous) blood of the coronary system of the beating heart *in situ*, and found that 0.39 mg of glucose and 3.1 mg of lactic acid per gram of heart tissue per hour disappeared. They concluded that glycolysis alone accounted for the loss of glucose, none of which was actually absorbed by the heart muscle. Comparable results were obtained by Evans and his associates in their earlier experiments, but in a more recent study they found that glucose was also consumed by the heart muscle though to a much less extent than was lactic acid. According to these workers, the blood lactic acid of the intact resting animal is derived from the breakdown of glucose in the lungs and in the blood itself. The lactic acid is oxidized by the heart muscle and other tissues. During strenuous muscular exertion lactic acid is produced in large quantities in the active muscles and passing into the blood is removed in the same way (ch. 52). The lactic acid utilization by the heart rises as its work increases. The glucose removed is used mainly to restore the glycogen stores. Bogue and his associates found that the heart's glycogen store is rapidly depleted by the administration of adrenaline. It is also depleted in the isolated heart perfused with saline (lactate unavailable), or in a heart inadequately supplied with oxygen. The disappearance of glycogen is due to its conversion to lactic acid. On the other hand, if the oxygen supply is plentiful and lactic acid is available it is utilized, and the glycogen stores conserved. The glycogen stores can be replenished by glucose but not by lactic acid. Thus lactic acid while it can prevent depletion of cardiac glycogen cannot replenish the glycogen stores. These facts point to glycogen as an emergency fuel rather than one which is utilized under ordinary circumstances.

The chemical changes occurring in cardiac muscle during activity, though resembling those taking place in skeletal muscle, differ from them in certain important particulars. In the first place, as just indicated, the oxidation of lactic acid occupies a much more important place in cardiac than in skeletal muscle contraction.

Cardiac muscle contains adenosinetriphosphate and phosphocreatine as does skeletal muscle,⁴ and as one would expect the breakdown of the former probably furnishes the brief burst of energy for the contraction, the energy for resynthesis is provided by the oxidation of lactic acid. In the absence of oxygen glycogen is broken down, from which reaction, it is believed, energy is then derived for the resynthesis of adenosinetriphosphate and phosphocreatine. When poisoned with iodoacetic acid cardiac and skeletal muscle behave differently. With an adequate supply of oxygen the poisoned heart muscle continues to contract indefinitely, its utilization of lactic acid is not impaired, and neither adenosinetriphosphate nor phosphocreatine are reduced. But in the absence of oxygen, lactic acid is not produced from glycogen and lactic acid available from any other source cannot be oxidized. Reduction of the phosphocreatine store to about half occurs and of the adenosinetriphosphate to about three quarters, the heart fails after a few beats (see Clarke and associates).

The heart muscle can also utilize fatty acids drawn from the blood or fat in its own substance. Fletcher and Waters, employing the heart-lung preparation perfused with heparinized blood containing no glucose (washed corpuscles in plasma previously fermented with baker's yeast) and a very low lactic acid concentration, found that in a heart performing moderate work over a 2 hour period no demonstrable reduction of the glycogen stores occurred, nor was there any diminution in the concentration of blood fat. From the results of other workers (Cruckshank and McClure) it appears that the mammalian heart does not utilize amino-acids as a source of energy to any important extent. The results of Fletcher and Waters' experiments indicate therefore that the heart can derive its sole energy requirements when necessary from noncarbohydrate material, most probably fat, composing its substance. These workers also suggest that glycogen is used only as an emergency fuel, that is, when heavy work is undertaken. Evidence for the utilization of fat by the cardiac muscle has been furnished by the work of Cruckshank and McClure and by Visscher, and of blood fatty acids by Cruckshank and Kosterlitz. Under certain conditions only a small fraction of the oxygen consumed by the heart can be accounted for by the oxidation of carbohydrate food (lactic acid and glucose), the balance being used presumably in the combustion of fat. On the other hand, with a high blood sugar the respiratory quotient (p. 374) may be 0.95 or 1.0, indicating a high carbohydrate utilization. The conclusion to be drawn from the results of the work of the various investigators in this field, is that the heart muscle shows a great adaptability in the utilization of different food materials. When lactic acid, glucose and fat are available all three may serve as sources of energy.

⁴ It also contains actomyosin indistinguishable from that present in skeletal muscle.

THE ORIGIN AND MODE OF TRANSMISSION OF THE HEART BEAT

If the heart of the frog is watched as it beats, waves of contraction may be observed to commence in the region of the sinus venosus and to pass over the auricle to the ventricle in orderly sequence. This wave of contraction is referred to as the beat of the heart. We know that the visible contraction is preceded, by a measurable interval of time, by an electrical change. Also, physico-chemical changes whose precise nature is unknown, undoubtedly precede or accompany the changes in electrical potential. These electrical and chemical changes which sweep over the heart in advance of the mechanical change are spoken of as the *cardiac impulse* or *excitation wave*.

HISTORICAL. Though Harvey (1628) studied the movements of the heart and discovered the circulation of the blood, he offered no explanation for the origin of the beat. Lower (1631-1691) thought that the beat was initiated by animal spirits which were supplied to the cardiac muscle through the vagus nerves. In the eighteenth century Haller, who observed that the spread of the contraction and the flow of blood through the chambers of the heart were events which proceeded hand in hand, concluded that the contact of the blood with the cardiac tissue at successive points was the direct stimulus for the muscular contraction. The contraction was looked upon as a simple peristaltic wave. Haller also pointed out that the right auricle was the last part of the heart to cease beating—the *ultimum moriens*—a fact consonant with his theory, since this chamber usually contained more blood at the time of death than other regions of the heart.

The latter view concerning the cause of the heart-beat was held until it was found that the excised heart would beat though its chambers were quite empty. It was then realized that whatever started the beat must arise within the heart itself. But other questions arose. Was the beat initiated by nervous tissue (ganglion cells) within the heart and carried to all its parts by nerve fibers or was the power of rhythmical contraction inherent in the muscle fiber? Two opinions have been held as to how these questions should be answered. One school, the *myogenic*, adhered to the view that the muscle fiber itself initiated the impulse. The other, the *neurogenic*, maintained that the impulse had its origin in nerve cells scattered among the muscle fibers and was conveyed for shorter or longer distances by nerve filaments. Support was given to the latter view by the

discovery of three groups of ganglion cells in the frog's heart which were placed at what appeared to be strategic points for the control of the beat. These groups of cells are known respectively as the *ganglia of Remak, Ludwig and Bidder*. The ganglion of Remak lies in the sinus venosus at its junction with the right auricle. Ludwig's ganglion is situated in the interauricular wall and Bidder's in the auriculo-ventricular groove. According to the neurogenic hypothesis Remak's ganglion was a nerve center which sent motor impulses to the cardiac muscle much as the respiratory center emits automatically a stream of impulses to the diaphragm and the other muscles of respiration. The other two centers were considered more or less subsidiary to this main one and to aid in the coordination of the beat.

THE FIRST AND SECOND STANNIUS LIGATURES Stannius (1852) showed that if a ligature were tied tightly around the heart of the frog or turtle at the junction of the sinus with the auricle, the heart below the ligature ceased to beat though the rhythm of the sinus remained unaltered. After a time the auricles and ventricles recommenced to beat but at a different rate than the sinus. If while the ventricle was under the influence of the first ligature a second one were tied around the auriculo-ventricular groove the quiescent ventricle at once commenced again to beat. From the neurogenic point of view the following explanation seemed quite satisfactory. Stoppage of the auricles (atria) and ventricles by the first ligature was due to the injury sustained by the chief motor center—Remak's ganglion—and the removal of its influence. The second ligature re-established the ventricular beat by stimulating Bidder's ganglion. Yet an explanation based upon the myogenic conception was equally plausible. Impulses arising in the sinus musculature it was claimed were blocked by the compression of the muscle fibers by the first ligature. The auricular and ventricular muscles after a time resumed their beats spontaneously by virtue of their inherent power. The second ligature restored the beat by direct stimulation of the muscle in the auriculo-ventricular groove.

Gaskell (1881) showed that compression of the cardiac tissue not only at the sino-auricular junction, but anywhere between sinus and ventricle, as may be accomplished by a specially devised clamp (Gaskell's clamp), would cause standstill of the heart below the area of compression. This is so though the tissue between the jaws of the clamp is devoid of nervous elements. The part below the clamp later commenced

to beat at a rate of its own. Gaskell also found that gradual compression of the tissue in the auriculo-ventricular groove stopped the contractions of the ventricle. If the ganglion cells in this region (Bidder's) were then exposed and stimulated by a needle no contractions were called forth. On the other hand, when the muscular tissue, free from nerve cells, was pricked rhythmical beats followed. It has also been shown that the tissue of the great veins which contain no ganglion cells exhibits spontaneous rhythmicity when completely separated from the sinus. Furthermore, in the heart of the tortoise it is possible to separate the sinus and the ventricle, except for a narrow bridge of muscular tissue and a single nerve (the coronary nerve). The only nervous communication between the two chambers is by means of this nerve. If in a heart so prepared the coronary nerve is cut, no change in the rhythms of sinus and ventricle occurs. On the other hand, if the nerve trunk be left intact but the muscular connection severed, the two chambers then beat quite independently of one another.

From the results of these experiments it was difficult to draw any other conclusion than that the beat of the primitive heart was myogenic. Gaskell also showed that muscular tissue in different regions of the heart possessed rhythmical powers of varying degrees. The sinus was shown to possess the greatest rhythmical power, the auricle in the region of the auriculo-ventricular junction was endowed with less and the ventricle with the least. The sinus, it was believed, originated the beat under ordinary circumstances, from here the impulse spread through the muscular tissue to the other regions of the heart. Though the sinus usually dominated the rhythm of the heart another region could, upon occasions when the function of the sinus was depressed, usurp control and develop a beat of its own.

Apparently there was no obstacle to the acceptance of the myogenic conception in so far as the amphibian or the reptilian heart was concerned, since in them the musculatures of the auricles and ventricles are continuous across the auriculo-ventricular groove. But the muscle tissue of the upper and lower chambers of the mammalian heart were apparently completely separated by the auriculo-(atrio)-ventricular ring of connective tissue. In the face of this fact there was no alternative but to believe that the transmission of the beat in higher animals was mediated by nerve fibers, and if this were so its origin in nervous tissue also must be granted.

The foregoing objection to the application of the myogenic theory to the mammalian heart was removed by the work of Kent who demonstrated the existence in rats of muscle fibers passing through the connective tissue of the A-V ring. Shortly afterwards His described in other mammals a definite bundle of primitive muscular tissue (bundle of His, or auriculo-ventricular bundle) connecting the right auricle and the

ventricles. This bundle and the nodes of similar tissue in the right auricle will be presently described.

In addition to the evidence just given the following facts in favor of the myogenic theory of the origin of the beat may be cited: (a) In the embryo chick the rudimentary heart shows rhythmical contractions after 36 hours. Ganglion cells do not appear until after the 6th day. The heart of the human fetus commences to beat after 3 weeks of gestation, nervous elements do not appear until 2 weeks later. (b) The muscular tissue of the apex is said to be free from ganglion cells yet a strip excised from this region beats rhythmically. (c) If zig-zag cuts be made in the ventricle so as to interrupt any nervous paths of transmission that might exist, the sequence of the beat is unaffected. (d) In several invertebrate hearts no ganglion cells can be demonstrated.

THE SPECIALIZED TISSUES OF THE MAMMALIAN HEART

The muscle fibers mentioned in the foregoing section which were discovered by Kent to form connections between the auricles and ventricles have subsequently been shown to belong to and form part of a special system for the origin and propagation of the beat in the mammalian heart. This system is composed of primitive muscle tissue interspersed with a few nerve cells and fibers. There is every reason to believe, however, that the initiation of the beat in, and its transmission through this system depends upon its muscular elements.¹

These specialized tissues comprise the following structures (fig. 20 1)

The sino-auricular node

The auriculo-ventricular node

The auriculo-ventricular bundle

The branches of the bundle and the Purkinje system

(1) *The sino-auricular or -atrial (S-A) node*

The sino-auricular (-atrial) node was discovered by Keith and Flack in 1907. It lies embedded in the muscle of the right auricle about the middle of a slight linear elevation situated on the inner surface of the chamber and known as the *taenia*

¹ The muscle cells have been shown to be capable of rhythmical contraction. Also if the continuity of the tissue be interrupted the gap is subsequently bridged over with scar tissue. Regeneration does not occur and conduction is never restored, this is contrary to what one would expect if the pathway of conduction were composed of nervous tissue. When the intestine, for instance, is cut across and the cut ends sutured together the divided fibers of Auerbach's plexus regenerate and functional continuity between the nervous elements in the two segments is soon restored.

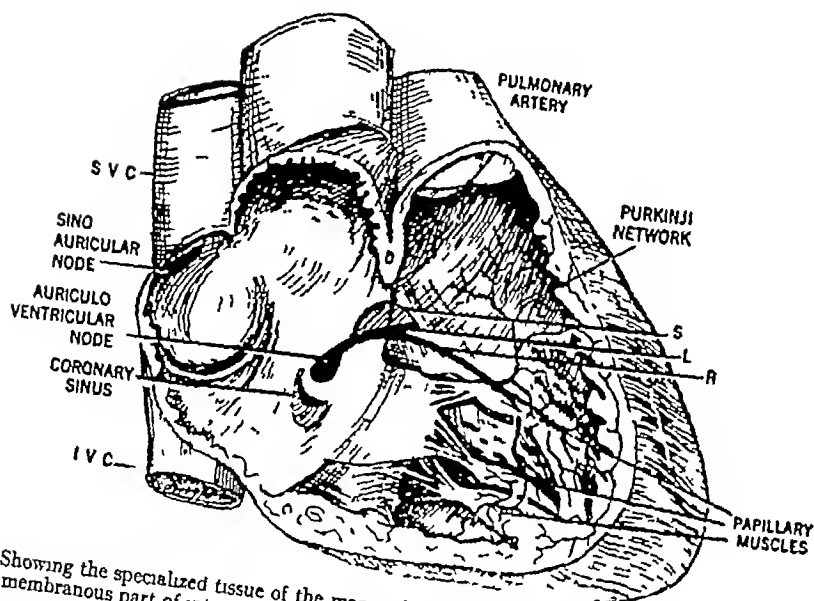


FIG 201 Showing the specialized tissue of the mammalian heart SVC, superior vena cava, IVC, inferior vena cava, S, membranous part of interventricular septum, L, left branch of bundle, R, right branch of bundle

terminalis The *tenia terminalis* corresponds to a shallow groove on the outer surface of the auricle—the *sulcus terminalis*—which runs downwards from a point to the right of the opening of the superior vena cava to the inferior caval opening. The node which is about $\frac{1}{2}$ inch in length has a special blood supply. Its minute structure is of a neuromuscular nature consisting of striated spindle shaped muscle fibers arranged in a plexiform manner, embedded in connective tissue and intermingled with a small number of ganglion cells and nerve fibers.

The muscle of the *tenia terminalis*, according to Prinzmetal, also serves during auricular systole to block the caval openings and prevent the reflux of blood into the great veins.

Knowing the ancestry of the node, which we shall see can be traced from the sinus tissue of the primitive heart, we may suspect its function. Just as the sinus initiates the beat and sets the pace for the cold blooded heart so the S-A node is the *pacemaker* of the mammalian heart. There exists for this statement, however, more than presumptive evidence based upon embryological considerations. The direct experimental evidence will be considered later.

(2) The auriculo (atrio-) ventricular (A-V) node

This was first described by 1906 by Tawara. It lies in the right auricle at the lower part of the

interauricular septum, anterior to the opening of the coronary sinus and above the septal leaf of the right auriculo ventricular valve. Its microscopical structure closely resembles that of the S-A node. Its power of impulse formation is apparently second only to that of the latter. When the S-A node is destroyed or its function depressed the A-V node may then assume the duties of *pacemaker*. A time interval of about 0.19 sec. elapses between the beginning of auricular and ventricular systoles of which 0.12 seconds represents a *delay* of the impulse at the node. The remaining 0.07 seconds is consumed in the passage of the impulse from the S-A to the A-V node. When the A-V node initiates the impulse the interval is much shorter or may be abolished, both chambers contracting nearly or quite simultaneously, since the impulse reaches both at about the same time. This is spoken of as *nodal* rhythm and occurs as a clinical disorder (p. 232). The S-A and A-V nodes are apparently unconnected by any tract of special tissue, the two being completely separated by auricular muscle. The primitive muscle cells of both nodes make intimate connection with the surrounding muscle fibers through the intermediary of cells which are transitional in structure between the cells of the node, and those of the auricular muscle proper.

The autonomic nerve supply of the S-A and A-V nodes is described in chapter 25.

(3) *The auriculo (atrio-)ventricular bundle, and (4) Bundle branches and Purkinje system*

It was mentioned on page 197 that following the discovery by Kent of muscular bridges connecting the auricle with the ventricle a well-defined bundle of muscular tissue was described by His. This bundle runs a short horizontal course forward and to the left from its origin at the A-V node, passing over the septal leaf of the tricuspid valve to the upper part of the interventricular septum. Here it divides into a right and left branch, each going to the corresponding ventricle. The ventricular divisions pass downwards beneath the endocardium of the septum and give off primary branches to the papillary muscles. The strands are continued beneath the endothelial lining of the ventricle and divide into innumerable filaments to form a delicate subendocardial interlacement. The muscle cells composing the ventricular branches and terminal arborizations of the system take on very special features, differing from those of the bundle stem and of the auriculo-ventricular node. The cells are much larger than in the latter situations, having a swollen appearance with large pale nuclei. These peculiar cells were described by Purkinje and had been known by his name long before their significance was recognized. Tawara's researches showed that such cells constituted the ventricular continuation of the A-V bundle. Nervous elements are also seen among the Purkinje fibers. The anatomical researches of Abramson and associates have shown that the Purkinje system does not consist simply of a layer of interlacing fibers beneath the endocardium, as has been supposed, but penetrates deeply into the ventricular muscle.

The auriculo-ventricular bundle and its branches are demarcated from the adjacent cardiac tissue by a fairly well-defined connective tissue sheath. Lhamon injected the sheath with colored fluids and found that the stem and branches were completely invested even to the finer terminations, also septa were found penetrating between the fibers to enclose individual Purkinje fasciculi.

THE DEVELOPMENT OF THE JUNCTIONAL TISSUES
The heart of lower vertebrates (frog, turtle, etc.) and the heart of the mammalian embryo possess an extra chamber—the *sinus venosus* (a, in fig. 20.2) which first receives the blood from the veins before it enters the auricle. In the early embryo two veins conveying blood chiefly from the upper part of the body and termed the right and left ducts of Cuvier open into the sinus. The blood passes from the latter through an

opening guarded by the *right and left venous valves* into a small ill-defined chamber, the *auricular canal* (b), which is continuous above with the auricle. Below where it forms a ring around the auriculo-ventricular junction, it is invaginated or telescoped into the ventricular cavity. This portion serves the purpose of a muscular bridge for the transmission of the impulse between the auricle and ventricle in the cold-blooded heart (p. 197). In the course of development of the mammalian heart these embryonic structures disappear but the tissue of which they are composed is believed to persist in vestigial form as the system of *specialized structures* already described, namely, the *sino-auricular node* and the *auriculo-ventricular node and bundle*.

The developmental history of these several parts of the system is as follows:

(a) The line of junction of the sinus venosus with the primitive auricle is represented by the *sulcus terminalis* of the adult heart. The tissue of the wall of the sinus at the opening of the right duct of Cuvier (which enters into the formation of the superior vena cava) and the tissue comprising the right venous valve are represented in the adult heart by the *sino-auricular node*.

(b) The sinus wall in the region of the left duct of Cuvier (which becomes in part the coronary sinus) and the tissue of the left venous valve remains as the *auriculo-ventricular node*.

(c) The funnel shaped invagination of the auricular canal is represented in the adult heart as the *auriculo-ventricular bundle*.

The investigations of Keith and Mackenzie suggest that the parts of the cold-blooded heart (sino-auricular and auriculo-ventricular junctions) from which the junctional tissues of the mammalian heart have appar-

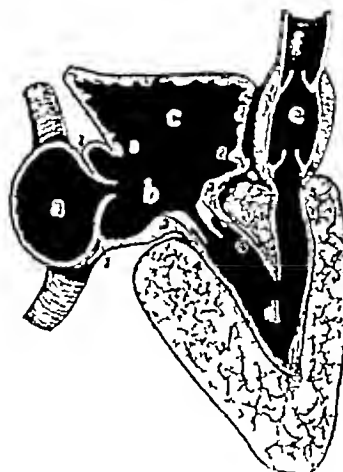


FIG. 20.2 A generalized type of vertebrate heart—combining features found in the eel, dogfish and frog (Keith), a, sinus venosus and veins, b, auricular canal, c, auricle, d, ventricle, e, bulbus cordis, f, aorta, 1-1, sino-auricular junction and venous valves, 2-2, canalo-auricular junction, 3-3, annular part of auricle, 4-4, invaginated part of auricle, 5, bulbo-ventricular junction, R D C and L D C, right and left ducts of Cuvier.

ently evolved are also of specialized neuromuscular structure (nodal tissue) They consider that this would explain the greater rhythmicity of these regions of the cold blooded heart The tissue is more diffuse in the hearts of lower orders, forming incomplete rings at the sino-auricular and auriculo-ventricular junctions, whereas in the mammal it is more localized This condensation becomes more pronounced as the animal scale is ascended

EVIDENCE FOR THE ORIGIN OF THE CARDIAC IMPULSE IN THE SINO AURICULAR (-ATRIAL) NODE

(1) Heat or cold applied locally to the node, but not to other parts of the heart, causes increase or decrease, respectively, in the rate of the heart beat

(2) Destruction or excision of the node causes the rest of the heart to stop beating for a time Such procedures are analogous to the first Stannius experiment as performed upon the cold-blooded heart Cohn, Kessel and Mason investigated the subject very fully in perfused hearts both inside and outside the body They found that excisions of portions of tissue in the region of the node merely accelerated the rate Incisions separating the node on three sides did not block the impulse The beat ceased, however, upon completion of the rectangle by a fourth incision The resumption of the beat in the auricles and ventricles was at a slower rate after the immediate effect of the incision (complete stoppage for from 4 seconds to 3 minutes) had passed off

(3) In a case of complete ectopia cordis in man, the impulse has been observed to originate in the region of the sino-auricular node, and to spread uniformly from there through the auricles The apex of the heart becomes excited slightly ahead of the base

(4) It was shown by Lewis that a contraction induced by artificial stimulation gave an electrocardiogram (ch 23) which most closely simulated the normal when the stimulus was applied in the vicinity of the S-A node itself Abnormal P waves appeared when any other part of the auricle was excited

(5) The crucial experiment proving the initiation of the beat in the S A node was devised by Lewis By means of a pair of contacts placed in different positions upon the surface of the normally beating heart of the dog, and connected with a galvanometer, Lewis demonstrated that at each beat the node was the first part of the heart to show relative negativity, which means that the tissue in this region was the first to become excited.

THE SPREAD OF THE EXCITATION WAVE THROUGH THE HEART

The transmission of the excitation wave from the sino-auricular (-atrial) to the auriculo (atrio-) ventricular node

The impulse is transmitted through the muscular tissue to the A-V node and across the inter-auricular septum to the left auricle It spreads at practically equal rates in all directions, there is no tract of specialized tissue offering a pathway of "least resistance" to the auriculo-ventricular node On this account it is difficult to produce a complete auricular blockage of the impulse The practical uniformity of the transmission rates along different paths radiating up or down to the right or left from the S-A node has been clearly demonstrated by Lewis Contacts were placed in series upon the surface of the auricle or upon the great veins and the arrival of the impulse at different points accurately timed The spread upwards through the wall of the superior vena cava was found to be an exception to the general rule, the rate being somewhat slower than in other directions Lewis has shown that the equality of the transmission rates through the auricle has an anatomical basis The S A node holds a strategic position in relation to the auricular musculature The chief muscle bands of the right auricle radiate from the region immediately below the node, termed the *concentration point*, like the sticks of a Japanese fan The average rate of propagation of the excitation wave through the auricular muscle is 1000 mm per second

The transmission of the impulse from auricle to ventricle

The auriculo-ventricular node, which we have already seen forms intimate connections with the auricular fibers, forms the first link in the chain which transmits the impulse to the ventricle The bundle, its right and left branches and the terminal arborizations of the latter constitute the rest of the transmission system *In the normal mammalian heart the bundle is the sole path by which the cardiac impulse can be conveyed from auricle to ventricle* Several investigators have demonstrated this fact conclusively by simply cutting or crushing the narrow bridge of special tissue This is invariably followed by complete dissociation of auricular and ventricular rhythms The cardiac disorder produced in this way is spoken of as *complete experimental heart block*

Erlanger and Blackman crushed the bundle in dogs under aseptic conditions by a specially devised clamp (Erlanger clamp). The animals lived from 320 to 340 days and during this time showed complete heart block. These experimenters showed what had not been shown before, that no other tissue could take over the duties of the bundle though the period of survival was sufficient, were it possible for such a vicarious action to be exercised. The experiment also shows that regeneration of the bundle does not take place. Epileptiform and syncopal attacks were observed in some of the animals. This is of much interest in view of the occurrence of similar manifestations in man as a result of disease of the bundle (see heart block, p 227).

The two ventricles do not contract exactly simultaneously. Sometimes one, sometimes the other chamber precedes its fellow by one or two hundredths of a second. The interval between the excitation of the auricles and the ventricles (A-V interval) has a duration in the human heart of between 0.12 and 0.20 seconds. The most widely accepted theory of this delay in the spread of excitation to the ventricles is the slower conduction rate of the tissue composing the A-V node. Others attribute it to a delay in the passage from auricular muscle to the nodal tissue, as a result of a latent period of the latter. Another view is that the delay is due to the latency of the response of the ventricular muscle or of the A-V node to the auricular impulses. Abnormally long A-V intervals (over 0.20 seconds) constitute the condition known as heart block (p 227). Very rarely, shortening of the A-V interval to less than the lower normal limit occurs. It is due, in all probability, to the existence of an aberrant or accessory auriculo-ventricular bundle.

The spread of the excitation wave in the ventricle

The transmission of the impulse to the various regions of the ventricular musculature, right and left, is entirely by way of the His-Tawara system. Just as compression or section of the stem of the bundle causes complete functional dissociation of auricle and ventricle, so injury to one of the main (septal) branches interferes with the passage of impulses to the corresponding ventricle (p 225, 229). The view is now widely accepted that the excitation wave after traversing the main branches follows the fine arborizations of the Purkinje pathway lying beneath the endocardium, and so

completes a semicircular journey in each ventricle from the upper part of the interventricular system to the base of the heart. This takes 0.04 second and is at a rate of about 5000 mm per second.

By means of a series of contacts placed upon the surface of the ventricle, Lewis and Rothschild timed the arrival of the impulse at successive points. It was found that the wave did not follow the course of certain muscular bands as might be expected if the transmission were muscular. For instance, a well-marked band is found in the dog's heart which, commencing near the origin of the pulmonary artery, sweeps to the left across the interventricular groove and around the left border of the heart to the apex. This band which is of considerable length, was not found to be excited progressively from above downwards, i.e., from the base of the heart to the apex, but almost simultaneously throughout its entire length. The times of arrival at different points were 0.0241, 0.0231, 0.0198, 0.0150, 0.0146, 0.0187 and 0.0196 second (fig 20.3).

If the entire surface of the ventricle is investigated in this way it is found that the order of spread cannot be explained by the disposition of the muscular fibers. The wave appears first of all upon the surface of the right ventricular muscle in the vicinity of the interventricular groove which

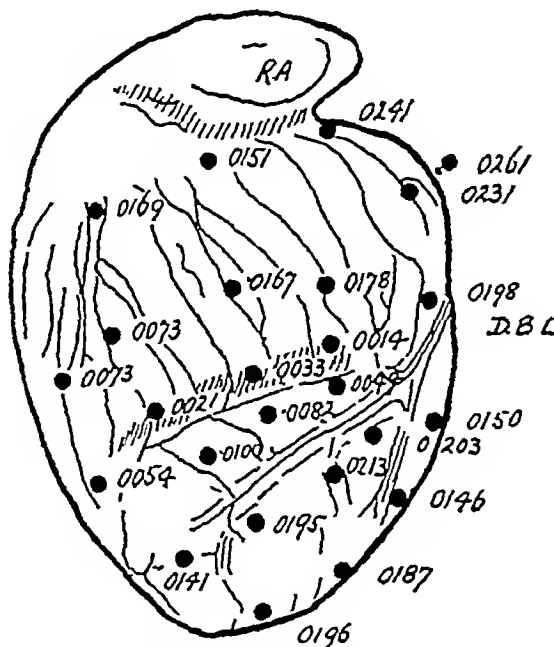


FIG 20.3 The times at which the excitation wave appears on the anterior surface of the heart (After Lewis)

may be shown to overlie the right papillary muscle and the early arborization of the right bundle branch (arrival times 0 0021, 0 0033, 0 0014 shown in figure 20 3) The wave then spreads to the rest of the right ventricle, the muscle at the base being the last to be excited In the left ventricle the apex is activated first and the mass of the muscle later The wave appears last at the base

Further evidence that the wave is conducted through the subendothelial system and not through the muscle is furnished by the following experiment. If two contacts (A and B, fig 20 4) be placed upon the surface of the ventricle and a mere scratch be made upon its interior so as to injure the Purkinje network lying between the contacts, then, when the ventricle is stimulated at the point shown in the figure, the arrival of the excitation wave at the upper contact is very markedly delayed The rate of propagation falls to that characteristic of ventricular muscle, namely, 500

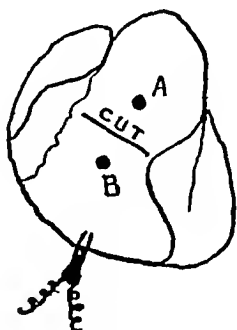


FIG 20 4 Description in text. (After Lewis)

mm or so per second On the other hand, if a deep cut be made from the outer surface through the muscle lying between the contacts, but the Purkinje system is not injured, no change in the transmission rate occurs

Abramson and others have also decided that the wave reaches the individual muscle fibers via the myocardial extensions of the Purkinje system (p 197) The muscle fibers in numerous places are thus excited almost simultaneously The terminal arborizations of the Purkinje system penetrate the muscle from endocardial to epicardial surface, making connection with the muscular tissue proper through the medium of transitional cells

The facts detailed in this chapter concerning the origin and spread of the excitation wave may now be summed up The beat is initiated in the tissue of the sino-auricular node From here it is "broadcast" through the auricular musculature with practically equal velocity in all directions The A-V node acting, so to speak, as a relay station, receives the impulse and transmits it to the ventricle via the auriculo-ventricular bundle At the upper part of the interventricular system the pathway forks and the wave of excitation passes down the septum on either side beneath the endocardium then to the muscle in the region of the interventricular grooves and lateral walls of the ventricles The evidence seems conclusive that the mode of spread of the impulse to all parts of the cardiac muscle is through the ramifications of the Purkinje network, and that the impulse does not follow the course taken by the muscle bundles

CHAPTER 21

THE ARCHITECTURE OF THE HEART, MOVEMENTS OF THE HEART, THE CARDIAC CYCLE

THE ARCHITECTURE OF THE HEART

The heart is irregularly conical in form. Its ventricular part constitutes a two-"cylinder" pump with muscular walls. The apex, composed entirely of the left ventricle, is directed forward, downward and to the left while the base, formed by the auricles (atria), and almost entirely by the left auricle looks backward. The ventricular cavities are enclosed by the lateral walls of the ventricles, the interventricular septum, which separates the two cavities, and the auriculo-ventricular septum. The walls of the auricles are relatively thin and weak, and are composed of a superficial and a deep layer of muscle. The former consists of fasciculi, which running transversely embraces both auricles, the deep layer of looped and annular fibers surrounds each chamber separately. The lateral ventricular walls, especially the wall of the left ventricle, are composed of thick and powerful muscle bundles. These are disposed in a complicated pattern of loops and whorls, which surround each ventricle separately or both together. At the apex they make a number of turns to form a sort of vortex and plunge deeply into the ventricular substance and then turn toward the base in the interventricular septum, some of these go to form the papillary muscles.

The pericardium. The heart is enclosed within the *pericardial sac* which is really two sacs, one within the other. The inner sac or *serous pericardium* consists of a single stratum of flat endothelial cells forming a two-layered membrane. One layer (*-visceral*) is adherent to the surface of the heart and is known as the *epicardium*. The other layer (*parietal*) lines the outer sac or *fibrous pericardium* which is a tough, fibrous membrane. The two walls of the serous pericardium contain between them a thin film of serous fluid. Its epicardial layer extends above over the great vessels, from which it is reflected to form the parietal portion. The fibrous pericardium is continuous above with the outer coverings of the great vessels. Below, it is attached to the central tendon of the diaphragm, and in front to the sternum. Through the connections of the serous pericardium to the great vessels, the heart becomes attached indirectly to the root of the lung, the diaphragm (especially the crura), and the structures at the root of the neck. From

these the heart obtains fixed points, or fulcra, to execute its movements. Otherwise the heart moves freely within the fibrous pericardium, and the parietal layer of the serous pericardium.

Besides anchoring the heart and so providing fulcra for its action, the pericardium serves two other functions to afford a smooth, lubricated surface for the heart's movements, and to serve as a check to overstretching of the cardiac fibers which otherwise might result from an excessive venous inflow during diastole.

THE MOVEMENTS OF THE HEART—THE APEX BEAT

The auricles exert little force in propelling the blood into the ventricles. They serve chiefly as reservoirs, their capacity for this purpose in the human heart being somewhat greater than that of the ventricles. Owing to the disposition of the muscle bundles, the blood within the ventricular cavities is propelled into the arteries by a sort of combined squeezing and wringing action of the ventricular walls. A movement equal in importance to the contraction of the lateral walls is the descent of the auriculo-ventricular floor. This is brought about by the shortening of the interventricular septum, which is the first part of the ventricle to contract. Shortly after, the papillary muscles contract and through the chordae tendineae acting as guy ropes support the auriculo-ventricular valves.

When the ventricles contract, all their diameters are shortened, but no change in the level of the apex occurs. When beads of opaque material are sewn along the line of the auriculo-ventricular groove of an animal's heart, and the chest after closure is viewed by X-rays, this region is seen to move downward during systole and up again during diastole. The auricles and great vessels become elongated, they cannot move downwards as a whole since they are fixed above. At the same time the apex rotates to the right, bringing more of the left ventricular surface to the front. In this movement the thick cardiac muscle, especially in the apical region, presses sharply against the chest between the 5th and 6th ribs in the mid-clavicular line, i.e., about an inch below the nipple,

producing an impact which can be felt by an examiner's fingers as a slight tap. This is called the apex beat. The slight movement of the skin of the 5th interspace can also be seen when the bare chest is viewed from one side in a good light, or it may be recorded by means of a tambour and writing point, such as are used in tracing the venous pulse (p. 226).

THE CARDIAC CYCLE

GENERAL DESCRIPTION

The succession of changes which occurs in the heart and is repeated during each beat is referred to as the *cardiac cycle*. On account of the rapidity with which the events in the cycle follow one another, it is impossible to study them by mere inspection. Harvey remarked upon the difficulties of the problem.¹ Modern methods of study include the graphic registration of pressure changes within the auricles and ventricles in animals, i.e., *intra-auricular* and *intra-ventricular pressure curves*. In man, records of the *arterial* and *venous pulses* (p. 226) obtained by means of the *polygraph* or of the electrical changes—*electrocardiograms*—(ch. 23) are employed in the study of cardiac action. It is also possible to record the pressures on the right side of the human heart by means of an intracardiac catheter inserted through an arm vein into the cardiac chambers (see ch. 26).

Before studying the intracardiac pressure curves and the methods by which they are obtained, a general account of the several phases of the cycle and their approximate time relations will be given. The contraction and relaxation of the auricles are called *auricular systole* and *auricular diastole* respectively. *Ventricular systole* and *ventricular diastole* refer to corresponding phases of the ventricular muscle. The length of the cardiac cycle,

when the heart is beating at the usual rate (70 per minute), is $\frac{4}{5}$ of a second (0.86 second). Its duration, of course, varies inversely with the heart rate. The moment in the cycle when ventricular systole has just come to an end may be chosen as the most convenient starting point for a description of the succession of events. At this time, the auricles and ventricles are relaxed. Blood is pouring into the cavity of the right auricle from the venae cavae and into the left auricle from the pulmonary veins. The auricular and ventricular cavities are separated from one another by the closed auriculo-ventricular valves. The semilunar valves also have been brought into apposition, so, during this period (0.08 to 0.12 second) the ventricles are completely closed but almost empty chambers. The auricles, on the other hand, are filled with blood which has accumulated during the previous ventricular contraction, but, as the pressure in the ventricle falls below the intra-auricular pressure, the auriculo-ventricular valves are opened, venous blood under a pressure of a few millimeters of water then pours into the fully relaxed ventricles. The latter chambers continue to fill throughout the remainder of diastole, that is, for about 0.40 second, nearly half the total length of the cardiac cycle. Toward the end of ventricular filling, however, the auricle contracts and the flow into the ventricle is hastened. The right auricle contracts slightly (0.013 sec.) before the left. Auricular systole lasts for 0.1 second. As a rule no interval or a very short one (0.016 second) elapses between the end of auricular and the commencement of ventricular systole. The fall in pressure in the auricle consequent upon its relaxation, and the rise of pressure within the ventricle as it contracts closes the A-V valves. The ventricle is again for a time (0.04 to 0.06 second) a closed cavity, since the semilunar valves have not yet opened. During this period the intra-ventricular pressure rises rapidly. The semilunar valves are forced open and the blood is discharged into the aorta and the pulmonary artery. The pressure continues to rise but soon reaches a maximum and is maintained around this level for a short time. The ventricle then enters upon its relaxation phase, the pressure falls and the semilunar valves close again. This brings us around to the point from which we started, a cardiac cycle has been completed.

Emphasis should be laid upon the following points in the foregoing description:

(a) Blood pours into the auricles throughout

¹ "When I first tried animal experimentation for the purpose of discovering the motions and functions of the heart by actual inspection and not by other people's books, I found it so truly difficult that I almost believed with Fracastorius that the motion of the heart was to be understood by God alone. I could not really tell when systole or diastole took place, or when or where dilatation or constriction occurred, because of the quickness of the movement. In many animals this takes place in the twinkling of an eye, like a flash of lightning. Systole seemed now here, now there, diastole the same, then all reversed, varied and confused. So I could reach no decision, neither about what I might conclude myself nor believe from others. I did not marvel that Andreas Laurentius wrote that the motion of the heart was as perplexing as the flux and reflux of Eurypus was to Aristotle." William Harvey, "Exercitatio anatomica de motu cordis et sanguinis in animalibus," 1628.

the entire cycle except for the short period (0.1 second) occupied by auricular systole

(b) Filling of the ventricle continues throughout the cycle except during its contraction and the brief succeeding period (0.08 second) which intervenes between the closure of the semilunar and the opening of the A-V valves

(c) The ventricle contracts for about 0.30 second and rests for about 0.50 second. The auricle contracts for 0.1 second and rests for 0.7 second. The resting periods of the two chambers therefore overlap for a period (about 0.40 second) during which the whole heart is quiescent

The following table gives the approximate figures for the duration of the chief phases of the cardiac cycle when the heart is beating at the usual rate of 70

	Seconds
Ventricular systole	0.3
Ventricular diastole	0.5
Auricular systole	0.1
Auricular diastole	0.7
Quiescent period of whole heart	0.4

THE GRAPHIC REGISTRATION OF INTRA-CARDIAC PRESSURES

The times of filling and emptying of the cardiac chambers, the pressures developed and the times of opening and closing of the valves guarding the orifices of the heart can be studied with great precision by this method

MANOMETERS The pressure changes within the heart cavities are recorded by means of optical manometers. A mercury manometer, owing to the great inertia of the column of metal, and the rapid changes in pressure which take place during the cardiac cycle, is unsuitable. The pressure in the left ventricle, for example, reaches a value of 100 mm Hg or more within a few hundredths of a second. On account of the lag of a mercury column the amplitude of the curve obtained by a mercury manometer is considerably less than it should be, and the recorded pressure is, in consequence, less than that which actually exists, also, minor fluctuations in pressure are not registered. Tambours covered with a thin rubber membrane, upon which a very light lever is fixed, have been employed with the view of overcoming these disadvantages. Marey's instrument is based upon this principle. In Hürtle's instrument, on the other hand, the pressure changes are made to bend a stiff spring, and in Fick's the curvature of a tubular spring is altered by pressure changes transmitted to its cavity. The moving parts of all these types, however, possess considerable mass and, though improvements over the mercury manometer, tend, nevertheless,

through their inertia to cause inaccuracies in recording. In the *optical method* an imponderable lever, namely, a beam of light, is employed. The movements of the beam of light are recorded photographically. This and the other methods just referred to must, of course, be calibrated against a mercury column in order to obtain quantitative results. A description of Wigger's modification of Frank's optical method follows.

The instrument (fig. 21 1, I and II) consists of (1) a *glass tube* which is introduced into the ventricle, the auricle or the aorta. The tube is surmounted by a brass chamber (c) which can be connected or disconnected with the lower part of the tube by means of a stopcock (g), and the pressure thus transmitted to the recording part of the apparatus as required. (2) A segment capsule (b) carrying a small mirror closes the brass chamber above, (3) a *moving photographic surface*—plate or film.

The segment capsule (fig. 21 1, III) consists of a small chamber 3 mm in diameter and circular in cross section except that a segment with a small arc has been cut off one side. The chamber is covered with a tightly stretched, thin rubber membrane. A small trapezoid celluloid plate is cemented to the surface of the membrane with the longer of its unequal sides lying flush with the chord of the capsule, while its shorter side does not extend as far as the center of the circle. The mirror is cemented in turn to the celluloid plate with its center lying precisely over the chord of the capsule upon which it pivots. By means of this device, changes in pressure transmitted to the membrane cause movements of the mirror which are greater than those of the membrane itself. The instrument is firmly clamped in position and a strong light focused upon the mirror. The beam is reflected to a second larger mirror which is adjusted to bring the ray into the horizontal plane, and have it fall ultimately upon a photographic surface moving in a suitable camera (photokymograph) across the path of the beam. The longer the beam the greater, of course, will be the magnification, though there is a limit to the length (120 to 150 cm.) which can be used advantageously.

THE INTRA-AURICULAR PRESSURE CURVE

This shows three crests (positive waves) *a*, *c*, and *v* and three depressions (negative waves) *r*, *x'* and *y* (see fig. 21 2). The *first positive wave*, *a*, is caused by auricular systole. It rises fairly steeply to its summit but represents a pressure of only a few millimeters of mercury. It should be pointed out that all parts of the auricular muscle do not contract together but progressively with the spread of the excitation wave. During the rise in the *a* wave more and more fibers are entering into the excited state, tension is being developed and the auricular pressure is rising. At the summit of *a*

all fibers are contracting After this, the fibers relax also in a progressive fashion, those first excited being the first to relax, and the pressure falls The consequent decline in pressure produces the "negative" wave, x Wiggers refers to the time

during which a group of fibers remains in the contracted state as the *fractionate contraction* The shortening of the auricular muscle as a whole is the resultant of the contracted and the relaxed fractions Wiggers, therefore, divides auricular systole into two phases, a *dynamic phase*, in which the proportion of the muscle fibers in the contracted state is increasing and the pressure is rising to a maximum, and an *adynamic phase* in which the pressure is falling as a result of the fractionate contractions coming to an end in progressively greater numbers The duration of each phase is about 0.05 second

The *second positive wave*, c , is due to the rising pressure in the ventricle as this chamber commences its contraction The pressure is transmitted to the auricle through the closed A-V valves which are bulged into the upper chamber The summit of c coincides with the opening of the semilunar valves

The decline of the c wave and the production of the *negative wave*, x' , is ascribed to two factors (a) the increase in the negative pressure within the thorax resulting from the ejection of the ventricular contents—about 75 cc in man—from the thoracic cavity (p. 263), (b) the depression (drawing toward the ventricle) of the auriculo-ventricular septum as the ventricle contracts

The *third positive wave*, v , is a stasis wave It is due to the inflow of blood from the veins and its accumulation in the auricle while the A-V valves are closed It is therefore the result indirectly, as is the c wave, of ventricular systole

The decline of wave v and the production of the *negative fluctuation*, y , are due to the opening of the A-V valves and the escape of the blood into the relaxed ventricle A small notch which sometimes appears on the upstroke of the v wave near its summit is ascribed to a vibration set up by the closure of the semilunar valves

The pressure curves of the two auricles show similar features but are not perfectly synchronous As a result of the time taken for the excitation wave to spread to the left auricle, systole of this chamber commences 0.013 second after that of the right.

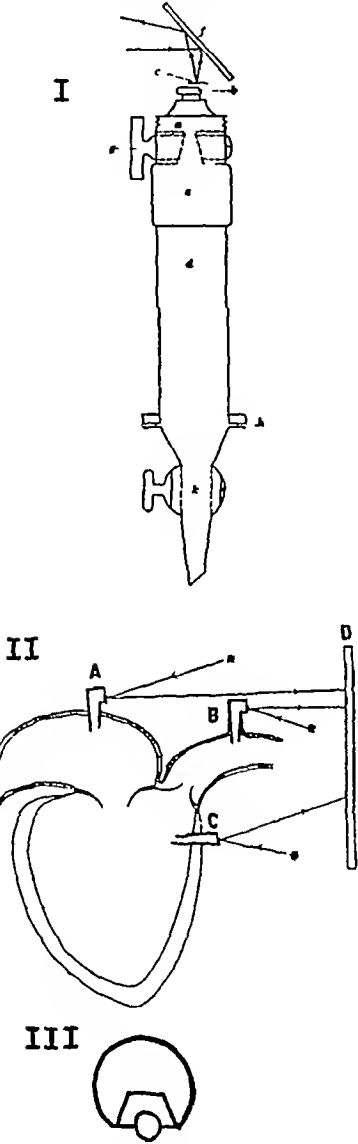


FIG. 21 1 I Optical manometer for registering intra-cardiac pressure a , damping plate, b , segment capsule, c , small Zeiss mirror, d , glass tube, e , brass cylinder, f , reflecting mirror, g , stopcock, h , conical joint for attaching different types of cannula II Diagram to illustrate optical method of recording A , manometer in auricle, B , in aorta, C , in ventricle, D , photographic surface. III Enlarged view of segment capsule, see text (after Wiggers, in part redrawn)



FIG. 21 2 Auricular pressure curve Description in text.

During ordinary conditions of bodily rest or light cardiac work the auricular contraction would appear to be of minor importance in the dynamics of the circulation, i e , in the filling of the ventricle. This must be evident when it is recalled that the dynamic phase of auricular systole occupies only 0.05 second while the total period of ventricular filling has a duration of about 0.4 second. Also in auricular fibrillation (p 235), in which condition the auricle can certainly exert no propulsive force upon the blood, filling of the ventricle is apparently not interfered with.

But according to Prinzmetal and his associates contractions of the right auricle and the right auricular appendix are forceful during heavy cardiac work due to increased venous return, and exert an important effect in filling the ventricles. The body of the left auricle, however, at no time serves otherwise than as a simple passage or conduit for the blood.

THE INTRA-VENTRICULAR PRESSURE CURVE

The auricle and ventricle throughout a large part of the cardiac cycle communicate with one another through the A-V opening, the pressures in the two cavities during this time are therefore virtually equal. When auricular and ventricular pressures are recorded simultaneously the two curves run parallel for long stretches (fig 21.3). A small positive wave is frequently seen just preceding the main pressure rise of the ventricular curve (line 8). This occurs synchronously with the *a* wave of the intra-auricular curve and is due to auricular systole. When the ventricle contracts (line 1) the pressure rises rapidly and almost instantly mounts above the auricular pressure. This causes the firm closure of the A-V valves which, as we have seen, actually bulge into the auricular cavity to produce the *c* wave. The pressure continues to rise and the steep upstroke of the curve is inscribed. Upon attaining a value in each ventricle just greater than the aortic or pulmonary diastolic pressure, respectively, the

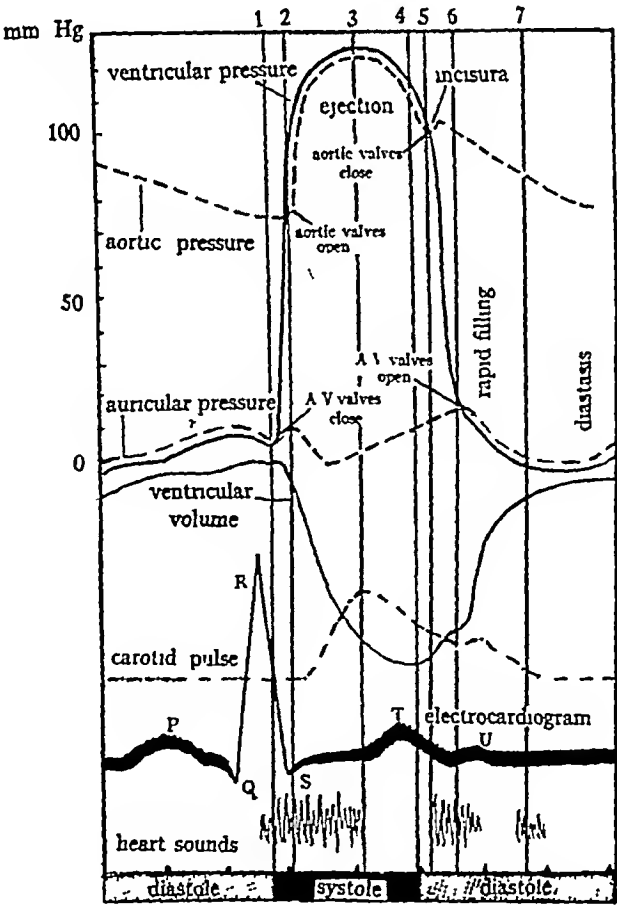


FIG 21.3 Superimposed curves of ventricular, auricular, and aortic pressure, together with a curve, an electrocardiogram, and a phonocardiogram (see text) Time, 1/10 sec

semilunar valves are forced open. This point is marked by line 2 in figure 21.3. The ventricular cavities and arterial lumina are now in free communication with one another, the ventricular and arterial pressures are about equal and continue to rise together for a time, but soon reaching their maxima, form the short plateaux or rounded summits of the curves. A sharp decline in the curve follows as systole comes to an end and the ventricular muscle commences to relax. The fall in intra-ventricular pressure below the aortic pressure causes the closure of the semilunar valves (line 5). This sharp drop in pressure at the termination of systole produces the incisura of the central arterial pulse (p. 187).

Following the closure of the semilunar valves and as a result of the more complete relaxation of the ventricle the pressure curve falls precipitously to reach finally a level below the auricular pressure (line 6). At this instant the A-V valves open and blood is again received from the upper chamber. Throughout the period which follows in which both auricle and ventricle are relaxed, pressures in the two cavities are approximately equal and at a low level. Toward the latter half of this period, however, the curves may show a gradual rise as a result of the distension of the heart cavities by the accumulated blood and the rise in venous pressure.

THE PHASES OF THE CARDIAC CYCLE AS SHOWN BY THE VENTRICULAR PRESSURE CURVE

The vertical lines numbered from 1 to 8 drawn through the ventricular curve indicate the occurrence of important events in the cardiac cycle. When, as in the figure, arterial, auricular and venous pulse tracings are accurately superimposed with the ventricular curve, that is, when all tracings commence at the same instant and are inscribed beneath one another, then the numbered lines will indicate synchronous events in the various curves at the points of intersection.

The phases of ventricular systole and the post-sphygmic period of ventricular diastole

Ventricular systole is represented by that part of the pressure curve included between lines 1 and 5. That is, from the closure of the A-V valves to the closure of the semilunar valves. In man while the body is at rest, the entire length of ventricular systole is about 0.30 second (from 0.25 to 0.36 second). Its length varies with the pulse rate, shortening with acceleration of the heart, and

vice versa. Systole is divided into two periods by the opening of the semilunar valves (line 2). The first of these phases (from line 1 to line 2) is termed the *presphygmic period* or *period of isometric contraction*. The former term implies that this period precedes the appearance of the arterial pulse (*G sphygmus*, pulse), the latter term connotes that the muscle fibers are not shortening. This is actually the case, since the fibers are at this time contracting upon a mass of liquid which is incompressible and completely fills a closed cavity. The second phase (line 2 to line 5) intervenes between the opening and closing of the semilunar valves and is termed the *sphygmic* or *ejection period*, since during this time the blood is being discharged into the aorta and pulmonary artery. The ventricles do not empty themselves completely during the ejection phase of the cardiac cycle, a variable amount of blood remaining in the heart at the end of systole. The short period of ventricular diastole immediately following the ejection period and up to the moment of opening of the A-V valves (line 6) is known as the *postsphygmic period* or *period of isometric relaxation*. During this period the ventricle is again closed and undergoing relaxation, no blood is entering its cavity so that the length of the fibers remains unchanged—hence the use of the term, isometric.

In man the presphygmic or isometric contraction period lasts for from 0.04 to 0.06 second. It is subject to little variation under changing conditions of heart rate, etc. The period of ejection takes up the balance of the time of the systolic period, i.e., about 0.25 second. Variations in the length of systole associated with changes in heart rate are brought about chiefly by changes in length of the ejection phase. This period is prolonged in aortic stenosis.

The commencement of the period of isometric contraction is synchronous with the initial vibrations of the first heart sound and follows by a very small fraction of a second the beginning of the R wave of the electrocardiogram (ch. 23). The termination of this period and the commencement of ejection is marked by the rise in the arterial pressure curve. The end of the ejection period coincides with the beginning of the second heart sound. Therefore, the total duration of systole may be determined in man from the interval elapsing between the initial vibrations of the two heart sounds (p. 210). The time from the commencement of the primary wave in the central

arterial pulse (p 187) to the bottom of the incisura will give the length of the ejection phase, i.e., the time elapsing between the opening and closing of the semilunar valves. The difference between the time of duration of the whole of systole and that of the ejection phase will give the length of the isometric period. An estimate may also be made of the total length of ventricular systole by measuring the interval from the beginning of R in the electrocardiogram to the end of T.

Wiggers takes as the end of the systolic period not the actual moment of closure of the semilunar valves but the point where the intraventricular pressure curve drops suddenly to produce the incisura. He further divides the ejection period into two phases. The first of these is termed the period of *maximum ejection* (from line 2 to line 3 in figure 21 3), which extends from the opening of the semilunar valves to the point of maximal pressure. The somewhat longer second phase during which the pressure is declining gradually and the outflow is lessening is termed the period of *reduced ejection* (line 3 to line 4). The interval elapsing between the end of the latter period and the closure of the semilunar valves, i.e., while the incisura of the central arterial pulse is being inscribed (line 4 to line 5), is included by Wiggers in the diastolic period and termed the *protodiastolic phase*. In the dog's heart the total duration of systole as thus defined was found to be from 0.15 to 0.26 second. The periods of maximum and reduced ejection were 0.05 to 0.12 and 0.06 to 0.14 second respectively.

Duration of different phases of ventricular systole in the human heart

Systolic phases	Total duration	seconds
Isometric contraction period (pre-sphygmic)	0.04-0.06	0.25-0.36
Period of ejection (sphygmic)	0.21-0.30	
Maximum ejection	0.09-0.14	
Reduced ejection	0.12-0.16	

The phases of ventricular diastole

The postsphygmic period or period of isometric relaxation and the protodiastolic phase of Wiggers have already been mentioned. After the opening of the A-V valves (at the end of the period of isometric relaxation) the blood flows rapidly into the ventricle, the intraventricular pressure curve

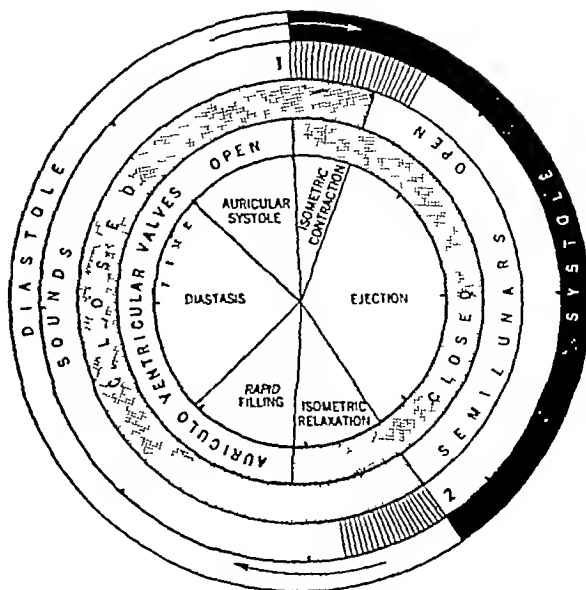


FIG 21 4 Diagram of cardiac cycle Time, $\frac{1}{10}$ sec

falls to the zero line (line 6 to line 7, fig 21 3). This is spoken of as the period of *rapid filling*. Unless the heart is beating rapidly the period of rapid filling is followed by one during which the ventricle, being already nearly full, fills much more slowly. During this period (line 7 to line 8) the ventricular volume increases very little, the intraventricular pressure may rise slightly. The term *diastasis* has been given to this phase by Henderson. Auricular systole follows the period of diastasis and, as already seen, causes the small pressure wave in the ventricular curve at the end of the diastolic period. It has been pointed out (p 206) that only the first half of auricular systole is responsible for this pressure rise. The period of diastasis is much curtailed or abolished when the heart beats rapidly, in which event auricular systole may ensue at the end of the phase of rapid filling, or after only a brief period of diastasis. The period of diastasis lengthens of course as the heart rate slows. The phases of the cardiac cycle are shown diagrammatically in figure 21 4.

Duration of different phases of diastole in the human heart

Diastolic phases	Total duration	seconds averages
Protodiastolic period (Wiggers)	0.04	
Isometric relaxation (post-sphygmic)	0.08	
Rapid inflow	0.09	
Diastasis	0.19	
Auricular systole	0.10	

CHAPTER 22

THE MOVEMENTS OF THE HEART VALVES THE HEART SOUNDS

The chief factor concerned in the opening and closing of the valves is, as already indicated (p 204), the difference in pressure upon their opposite surfaces. Some additional features of the valvular mechanisms must now be considered.

THE AURICULO-VENTRICULAR VALVES (TRICUSPID AND MITRAL)

The valve leaflets or cusps, three in number on the right and two on the left side, are attached by their bases to the fibrous rings surrounding the auriculo ventricular openings. Their free margins are connected through delicate tendons (chordae tendineae) to the papillary muscles which prevent inversion of the valves into the auricle during ventricular systole. The chordae tendineae are tightened at the commencement of systole by the contraction of the papillary muscles. The leaflets are composed mainly of a double layer of the endothelial lining of the heart, strengthened by a few connective tissue fibers. Their attached bases are thicker and contain more connective tissue, small blood vessels and delicate strands of smooth muscle. The latter, however, play no part in valve closure which is effected, as mentioned above, in a passive manner.

The mechanism of valve closure

During auricular systole the leaflets do not lie back against the ventricular wall, but occupy a mid position as a result of two opposing currents. The inflowing blood pressing upon their auricular surfaces keeps them open, while eddies reflected in the reverse direction from the ventricular walls strike their ventricular surfaces and tend to close them. Thus they float in a position of delicate balance. When, as a result of the fall in intra-auricular pressure at the end of auricular systole, the incoming jet is diminished in force and finally ceases, the back eddies persisting for a brief space and being unopposed, approximate the valves or bring them gently into apposition (fig 22 1). They are not, however, firmly closed. This is effected by the rise in pressure in the ventricle when it contracts. Dean has shown by attaching a hair

to the septal leaf of the valve and recording its movements, that if ventricular systole does not follow almost instantly upon the cessation of the flow of blood from the auricle the valves start to reopen. In instances, therefore, in which ventricular systole is delayed, that is, when the A-V interval is prolonged, the reopening of the valves proceeds for an appreciable time. Then, when ventricular contraction occurs, a small amount of blood regurgitates into the auricle before the valves are swung closed by the rising intraventricular pressure. The backward flow of blood through the orifice may then give rise to a murmur just preceding the first heart sound (presystolic murmur).

THE SEMILUNAR VALVES

The dynamics of aortic and pulmonary closure are essentially the same in principle as those described for the A-V valves. The valves form three small pockets open toward the arterial lumen. Back eddies which are set up during the ejection phase of systole prevent the contact of the valves with the arterial wall. When ejection ceases the centripetal currents carry the valves into apposition and firm closure is effected by the higher pressure at this time upon their arterial surfaces.

THE HEART SOUNDS

Two sounds can be heard during the cardiac cycle. The *first sound* is of relatively long duration, soft in quality and low in pitch. The *second sound* is shorter, sharper and of higher pitch. These characteristics are best imitated vocally by the syllables "lub" and "dup" separated by a brief pause. The two heart sounds mark the beginning and end of ventricular systole and the determination of the interval between their commencements (as determined by the phonocardiogram) is a reliable method for arriving at the length of ventricular systole in man. The pause between the end of the second sound and the beginning of the first coincides with ventricular diastole.

The factors concerned in the production of the sounds

I THE FIRST SOUND The first sound is heard most clearly and at maximum intensity over the fifth left intercostal space, i.e., within an area centered over the "apex beat" Here the mitral element of the sound predominates Any abnormal sound produced at the tricuspid valve is detected by listening over the lower end of the sternum

The principal factors entering into the production of the first sound are (1) the closure of the auriculoventricular valves and the tension set up in the valve leaflets and chordae tendineae as the intraventricular pressure rises (valvular element), (2) the rush of blood from the ventricles and the shock transmitted to the walls of the aorta and pulmonary artery (vascular element), and (3) contraction of the ventricular muscle Vibrations set up by the auricular muscle at the end of its systole may sometimes contribute to this sound

The vibrations of the valves is the most important element of the first sound Some have claimed that the first sound is not prolonged into the ejection period, and therefore the impact of the blood upon the walls of the large vessels cannot contribute to the sound But the observations of Straub and of Orías and Braun-Menéndez indicate that the first sound does extend beyond the isometric period and that the vascular element is an important component Of the three the muscular element appears to add least to the sound Dock has questioned whether the ventricular muscle produces a sound at all He recorded the sound vibrations by means of a phonocardiograph applied to the surface of the heart and reported that no sound was produced when the empty heart contracted, valve action being abolished Although he concluded from this that the normal first sound contained no muscular element, evidence from other sources is strongly opposed to such a conception Wiggers and Dean, previously, had recorded sound vibrations from an isolated strip of myocardium It has also been shown that when free movement of the valves of the beating heart is prevented, the booming character of the first sound still persists

Phonocardiographic records from the chest over the heart show, as a rule, no vibrations before the onset of ventricular contraction (fig 21 3) thus indicating that the presystolic apposition of the valves is silent The sound commences 0 008 second before the peak of the R wave of the electrocardio-

gram Its duration is about 0 18 second Sometimes, however, the phonocardiogram records a few small presystolic vibrations which are attributed to auricular systole, occasionally a faint sound is heard at this time

II THE SECOND SOUND results from the vibrations set up in the blood column and arterial walls, as the aortic and pulmonary valves are placed under tension following their closure The duration of the second sound is about 0 10 second It commences about 0 09 second after the summit of the T wave of the electrocardiogram It is heard best over the pulmonary and aortic areas—the upper part of the sternum adjacent to the second left intercostal space, and the second right costosternal junction, respectively

III THE THIRD HEART SOUND Sometimes a faint third sound is heard in normal hearts which follows the second sound by about 0 08 second and lasts for about 0 04 second It is heard at the apex and is commonly found in young adults Thayer found it present in 65 per cent of normal individuals It may be made to appear or is intensified by procedures which increase the venous flow into the auricles, e.g., exercise, recumbent position, etc The sound was first described by Gibson and independently by Herschfelder, several explanations have since been given to account for it. (1) Some have considered it to be due to the asynchronous closure of the aortic and semilunar valves, but the interval between it and the second sound is too great for the acceptance of this explanation (2) White considers that it is due most probably to the opening snap of the A-V valves or to the vibration of the ventricular walls as the blood rushes into the ventricle (3) Another view is that the sound is simply due to prolonged after-vibrations of the aortic valves which have become separated from the earlier vibrations as

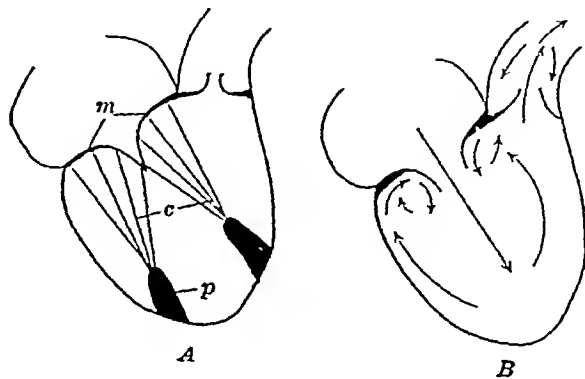


FIG 22 1 Two diagrams showing mechanisms concerned in valve closure A, showing relations of papillary muscles and chordae tendineae to valve flaps, B, partial closure due to eddy formation (after Wiggers)

a silent interval. None of these explanations is entirely satisfactory in all instances.

Variations in the intensity of the heart sounds

It is the general belief that the first sound varies in intensity with the force of ventricular systole and the loudness of the second sound with the height of the arterial blood pressure. The experiments of Wiggers bear this out. The sounds were recorded graphically and correlated with the intraventricular and aortic pressure curves. It was found that the vibrations of the first sound were increased in amplitude and number when the tension developed by the cardiac muscle was increased. The intensity of the first sound is directly related to the rate of the pressure rise within the ventricle during the isometric period. The intensity of this sound is not dependent upon the volume of the systolic discharge but rather upon the diastolic pressures in the pulmonary and systemic circuits. Wiggers found that when the heart was slowed and the systolic discharge consequently increased but the diastolic pressure lowered, the intensity of the first heart sound was reduced while acceleration of the heart (reduced

systolic discharge with raised diastolic pressure) increased its intensity.

The intensity of the second sound in the aortic or pulmonary area is increased by an elevation in the systemic or pulmonary pressures, respectively. Among the cardiovascular conditions associated with intensification of the second sound are, mitral stenosis and failure of the left ventricle which raise the pulmonary arterial pressure, and arterial hypertension which raises the aortic pressure.

THE GRAPHIC REGISTRATION OF THE HEART SOUNDS

The phonocardiogram The instrument employed for this purpose consists of a stethoscope fastened to the chest and provided with a side tube open to the air so as to obviate pressure changes, a microphone and a string galvanometer (fig 222). A type of apparatus devised more recently (Kauntz and associates), which they call the cathode ray "vibrocadiograph", avoids the acoustic errors of air transmission and employs a water filled capsule in contact with the chest. The vibrations are picked up by a dynamic type of microphone. The electrical currents are transmitted to an amplifier and thence to the vertical plates of two cathode tubes. The vibrations are recorded photographically.

The record of the first sound is composed chiefly of a series of from 9 to 13 vibrations, and has a duration of from 0.9 to 0.16 seconds. These are of small amplitude to start with but rise to a "crescendo" which reaches its maximum at the end of the isometric period and is followed by a "diminuendo" of about the same duration. This main series of vibrations is sometimes, as mentioned above, preceded by a couple of small introductory vibrations which occur prior to the ventricular pressure rise. These are possibly of auricular origin. The main series is also followed by a few final vibrations, variable in number. The vibrations are in general irregular, which places the sounds in the category of noises rather than of musical tones. The frequency is low, being on the average 45, 50, and 33 per second, respectively, for the first, second and third sounds. A normal and a series of abnormal phonocardiograms are shown in figure 223, page 213.

Graphic registration of the heart sounds has permitted precise relationships to be established between them and the events of the cardiac cycle, as recorded by other graphic methods, e.g., intracardiac pressure curves and the electrocardiogram. The phonocardiogram also, as already mentioned, provides an accurate method of measuring the length of ventricular systole.

The heart sounds may also be picked up by a microphone applied to the chest wall over the heart, amplified and conducted to a loud speaker. The fetal heart sounds can be amplified in this way so that they are clearly audible to the obstetrician during labor.

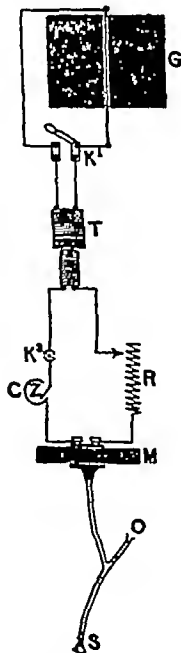


FIG 222 S, Stethoscope bell, O, side tube to air, M, microphone, CZ, dry cell, R, resistance, K₁, key in galvanometer circuit, T, induction coil, K₂, key in microphone circuit, G, galvanometer (after Lewis)

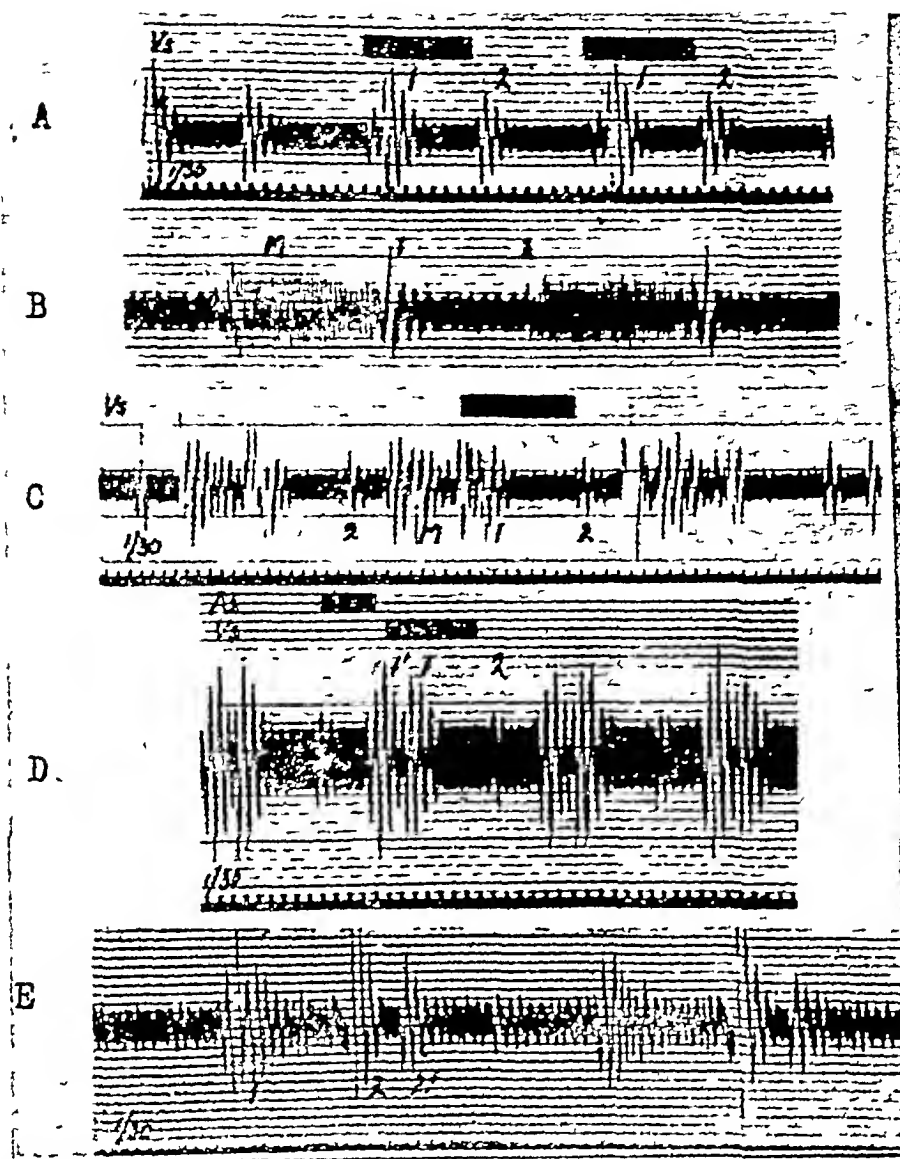


FIG 22-3 A, normal heart sounds recorded from apex, B, a diastolic murmur of aortic origin, C, presystolic murmur, D, gallop rhythm, presystolic (1') type, E, gallop rhythm, early diastolic (2') type. The numbers 1 and 2 refer to the heart sounds, M, murmur (after Lewis)

ABNORMAL HEART SOUNDS

Murmurs

When the valves at one or other of the cardiac orifices become deformed by disease, they impede the flow of blood or allow leakage to occur, and abnormal sounds replace either partially or completely the usual heart sounds. Such sounds are spoken of as murmurs or bruits. When the valves produce narrowing of one of the orifices of the heart, the condition is spoken of as stenosis, thus aortic or mitral stenosis implies restriction of the respective opening and increased resistance to the flow of blood through it. The velocity of the blood flow through the constricted orifice is also in-

creased. The valve surfaces are usually roughened and the smoothly rounded and somewhat funnel-shaped nature of the opening is lost. The normal stream-line character of the flow through the orifice is replaced by a turbulent flow. All these factors contribute to produce the abnormal sound. When the valves are incapable of closing tightly they are said to be incompetent. They no longer perform their duty but allow blood to pass through the orifice in a direction the reverse of that of the circulation. So aortic or mitral incompetence or regurgitation is spoken of to indicate leakage of the respective valve. Obviously if the valves are deformed sufficiently to produce narrowing of a particular orifice, they will also be incapable of

closing properly and leakage will occur as well. Stenosis is therefore usually associated with a certain degree of incompetence and *vice versa*

TIME RELATIONS The particular valve involved is determined from the relation of the murmur to the events of the cardiac cycle and from the point upon the chest wall where the sound is transmitted with the greatest intensity. For instance, the aortic valves should open fully during the ejection phase of ventricular systole so as to offer little or no resistance to the outflow of blood at this time. At the end of the period of ejection they should close tightly. If however the orifice is stenosed, the obstruction causes a murmur to be heard during ventricular systole which replaces or modifies the clear first sound. This is referred to as a *systolic murmur*. If, on the other hand, the valves are incompetent and do not come together at the end of the systolic discharge, a rush of blood from the aorta into the ventricle occurs in diastole, and a murmur modifies the normal second sound—*diastolic murmur* (fig 22.3 B). The murmur may appear in early, mid-, or late diastole or may persist throughout almost the entire period. When stenosis and incompetence co-exist a murmur may be produced at the aortic orifice during both systole and diastole, in which event the two normal sounds are replaced by a to and fro blowing sound. A systolic murmur will also be caused by incompetence of the A V valves (mitral or tricuspid) since the rise in pressure during ventricular systole will drive blood backwards into the auricle and cause abnormal vibrations to be set up.

Mitral stenosis (stenosis of the tricuspid orifice is very rare) by offering an obstruction to the flow of blood from the auricle into the ventricle may cause a murmur to be heard at any time between the second and the first heart sounds. The intensity, duration and timing of the sound vary with the degree of stenosis and the rate of blood flow through the orifice. Though the murmur may extend throughout diastole it tends to be more intense toward the beginning and the end of this phase, it may be heard only during these times, or be confined to one or other of them. As pointed out by Lewis, the rate of blood flow into the ventricle which is an important factor in the production of the abnormal sound, is most rapid during these phases of diastole.

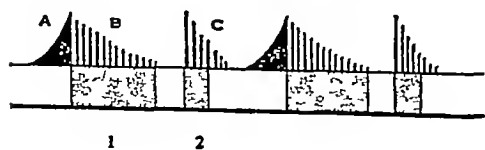


FIG 22.4 Diagram showing the time relations of heart murmurs to the heart sounds. A, presystolic murmur, B, systolic murmur, C, early diastolic murmur. Numbers refer to normal heart sounds

At the end of diastole an impetus is given to the blood by the systole of the auricle, while the early part of diastole is the period of rapid filling (p 209) of the ventricle. The late diastolic murmur (presystolic) may therefore disappear when auricular fibrillation supervenes and abolishes the propulsive action of auricular systole. When the heart is beating slowly the murmur is more likely to be well marked in early and in late diastole. The blood which has accumulated in the auricle during the preceding ventricular systole rushes through the mitral orifice when the valves open, but the flow slows as the ventricle becomes full (period of diastasis), so the murmur disappears. It reappears again when the auricle, distended with blood as a result of the prolongation of diastole, contracts. When the beat is rapid and the period of diastasis abolished, the murmur is likely to be heard throughout diastole which is then occupied entirely by the phase of rapid filling. The obstruction to the flow of blood through the narrowed mitral orifice causes a rise in auricular pressure which becomes transmitted back to the pulmonary veins and ultimately to the pulmonary arterial vessels and the right ventricle. Though the lungs contain a larger than normal amount of blood the flow through the vessels is slowed. Symptoms and signs referable to distention of the pulmonary capillary bed are prominent features, e.g., reduced vital capacity, dyspnea, cough, hemoptysis and cyanosis. The filling of the left ventricle during diastole is slowed and the auricle contracts with a greater force than usual at the end of this period. The cardiac output may be reduced, the left ventricle is normal in size or even smaller than normal while the left auricle and right ventricle are enlarged.

Aortic regurgitation causes a diastolic murmur, which may be heard only in early diastole or may persist into late diastole. It is due to the leakage into the ventricle through deformed aortic valves, which are not accurately apposed, their surfaces are usually roughened. The volume of regurgitated blood varies in amount, and the general signs vary with the extent of the leak. Since the deformed valves, as well as being incompetent, offer some obstruction to the outflow of blood from the ventricle (that is, a certain degree of stenosis of the orifice usually exists), a systolic murmur is often to be heard over the aortic area. A presystolic murmur is not uncommonly heard in aortic incompetence resembling that of mitral stenosis. This was described many years ago by Austin Flint and is usually referred to as the *Flint murmur*. According to the most generally accepted explanation, it is due to vibrations set up in the anterior leaflet of the mitral valve which is pushed by blood regurgitating through the aortic orifice into the path of the stream entering from the auricle. Gouley has described a characteristic deformity of the right leaflet of the aortic valve which he declares serves to direct the regurgitating blood against the mitral leaflet. Other features of insufficiency at the aortic orifice are (1)

great enlargement of the heart (*cor bovinum*) due chiefly to dilatation of the left ventricle and hypertrophy of its muscular wall. This chamber is called upon to accommodate in diastole not only the blood from the auricle but that regurgitated as well. And it must eject this greater volume of blood during systole. But the greater thickness of the ventricular muscle is not simply a work hypertrophy (see p 257), (2) low diastolic and high systolic pressure, abnormally high pulse pressure, (3) collapsing pulse, and (4) capillary pulsation.

The high initial tension at the end of diastole, which results from the filling of the ventricle by an unusually large amount of blood, causes a more powerful contraction. The isometric period is much shorter than usual and the pressure curve rises abruptly. As the semilunar valves open, an "explosive" ejection of the large volume of blood into the underfilled aorta causes the abrupt increase in systolic pressure. A much greater proportion than usual of the ventricular contents is discharged during the first half of systole, and a much smaller proportion than usual during the period of reduced ejection. It is at this time—the period of reduced ejection (p 209)—that the sharp fall in pressure occurs and *not during diastole*. The pressure fall continues steeply into the period of isometric relaxation. Little further decline during the remainder of diastole occurs. The leakage into the ventricle and peripheral vasodilatation are minor factors contributing to the low diastolic pressure characteristic of aortic regurgitation. The sudden ejection of a large volume of blood in early systole and of a relatively small amount in the second half of the ejection period also gives an explanation of the high-peaked character of the pulse tracing, and for the low position on the cataretic limb of the dicrotic wave, and also for the collapsing, or water hammer, character of the radial pulse (Corrigan's pulse). When the finger is held on the radial artery, a sharp impact is felt, followed by an equally sudden collapse, which is especially marked if the patient's arm is held vertically. A sound described as resembling a pistol shot may also be heard, when a stethoscope is applied over the artery.

The pulsation in the capillaries is attributed by Lewis and Drury to dilatation of the arterioles, which occurs for some unknown reason in aortic regurgitation (p 214), the arterial pulse being thus readily transmitted to the capillary bed. It can be seen most readily by slightly compressing the skin at the base of the finger nail. Wiggers found, however, that it could be produced in an artificial model, when central regurgitation was produced and the peripheral resistance kept constant. He therefore considers it to be due primarily to the low diastolic pressure and high pulse pressure, though its occurrence is undoubtedly facilitated by the dilatation of the arterioles.

Enlargement of the mitral orifice by incision of the

fibrous tissue surrounding it (mitral commissurotomy) is an operation practiced in recent years for the relief of mitral stenosis.

Aortic stenosis In severe stenosis of the aortic orifice, produced experimentally, the pressure in the left ventricle during its isometric contraction rises higher and more abruptly than in the normal heart, but both arterial systolic and diastolic pressures tend to be lower. The period of ejection is prolonged, this is largely responsible for the production of the typical pulse tracing shown on p 186. Owing to the narrowed aortic orifice, the blood is expelled from the ventricle at high velocity. The energy expenditure of the heart muscle for the work performed is increased (i.e., the mechanical efficiency is reduced). The velocity factor (p 142), instead of being around the normal of about 1 per cent of the total work, may constitute 10 per cent or more. The intensity of the systolic murmur increases with the degree of stenosis. The high pressure developed in the ventricle tends to be transmitted to auricle and pulmonary veins, and to the pulmonary arteries and right ventricle, when the left ventricle fails to maintain its normal output.

GALLOP OR CANTER RHYTHMS In certain cardiac conditions three distinct sounds are heard which give rise to a rhythm not unlike the gallop or canter of a horse. In some cases the abnormal sound precedes the first sound—the *presystolic type* of gallop rhythm. Two sounds are heard in rapid succession followed by a pause, and then by the second sound (fig 22.3, D). This type is associated with depressed auriculo-ventricular conduction (p 227) or sometimes with bundle branch block (p 229). The extra sound is attributed to the forceful contraction of the auricle. Normally the ventricular contraction follows so closely upon auricular systole that any sound that may be set up by the contraction of the auricle merges into the first sound. But when, as a result of slowed conduction in the A-V bundle, the auricular and ventricular systoles are separated by an appreciable interval the contraction of the auricular muscle may become audible. The unusually large load of blood retained in the auricle, due to the delay in the ventricular contraction, and the hypertrophy of the auricular muscle which frequently exists, also favor the production of the sound.

In other instances the abnormal sound follows shortly upon the second heart sound—*early diastolic type* of gallop rhythm (fig 22.3, E). In some instances the extra sound is simply an intensification of the third sound described above as occurring in normal subjects. It may then, though very occasionally, follow strenuous muscular exercise. In most instances this type of gallop rhythm is associated with cardiac failure and is then indicative, usually, of a severe myocardial damage. The abnormal sound is not caused by asynchronous closure of the pulmonary and aortic valves. It is

most probably due to vibrations set up in the walls of the dilated and relaxed ventricle caused by the shock of blood as it rushes from the auricle under a high head of pressure. The sound therefore follows upon the opening of the A-V valves, i e., after the period of isometric relaxation (p 209)

When the heart rate is slow and auriculo-ventricular conduction much prolonged, a type of gallop rhythm may develop in which the abnormal sound occurs near the middle of diastole—*mid-diastolic type* of gallop rhythm—and is attributed to a forceful contraction of the auricle.

CHAPTER 23

ELECTROCARDIOGRAPHY THE VENOUS PULSE

GENERAL DESCRIPTION OF THE ELECTROCARDIOGRAPH

The electrocardiograph as devised by Einthoven is, in essence, a very sensitive galvanometer. The cardiac action current is led through a fiber of finely spun silver-coated quartz glass, 0.002 mm in diameter (about $\frac{1}{4}$ of the diameter of a red cell). The fiber, or "string" as it is termed, is suspended vertically in a holder between the poles of an electromagnet. When the magnet is excited by a powerful current an electric field is set up, the lines of force passing from its north to its south pole. The heart current is conducted through the string (fig. 23.1). A strong beam of light is directed through apertures in the arms of the magnet lying in front of and behind the string. The latter and its lateral movements (deflections) are thus cast as shadows which are magnified and brought to a focus upon a moving photographic surface (plate, film or sensitive paper) moving vertically at the desired speed in a camera of special design.

The sensitivity of the string is standardized so that a deflection of 1 centimeter represents 1 millivolt. Horizontal lines a millimeter apart are marked on the record by means of etchings upon the camera lens. Each division represents $\frac{1}{5}$ millivolt (mV). The shadows thus cast upon the photographic surface are registered as the deflections or waves in the electrocardiogram mentioned above and described on pages 219 and 220.

The passage of the heart current causes circular lines of force to be set up around the fiber. These take a clockwise or anti-clockwise course according to the direction (descending or ascending) of the cardiac current, and cause the string to move, respectively, outwards or inwards toward the arch of the magnet. Time is indicated by vertical lines caused by some form of timing device which breaks the beam of light at regular intervals. Fine vertical lines mark time intervals of $\frac{1}{25}$ second, heavier ones of $\frac{1}{5}$ second.

The string galvanometer type of electrocardiograph is the one in most general use today but within recent years an instrument employing a moving magnet or moving coil galvanometer and radio tube amplification has been devised.

THE ELECTROCARDIOGRAPH LEADS

The standard leads We have already seen (p. 200) that the electromotive force developed in the cardiac muscle of an experimental animal can be recorded by placing paired contacts directly upon the surface of the heart and connecting them with a galvanometer or by placing one contact upon the heart and the other upon some indifferent part of the body. Records obtained in this way are called electrograms. It is clear that the electrical changes occurring immediately beneath a contact electrode will dominate the record. The heart *in situ* is, however, surrounded by a conducting medium—the blood and tissue fluids which are, in fact, solutions of electrolytes. It is possible, therefore, to connect the electric field set up about the heart during its beat by leading off from certain paired regions upon the body surface to the string of the electrocardiograph. A part of the electromotive force developed by the heart can in this way be "tapped" and the activity of any part of the cardiac muscle, and not merely the region beneath the electrodes as in the case of electrograms, will impress its influence upon the electrocardiogram. The parts of the body employed for this purpose are the two forearms (or hands) and the left leg.

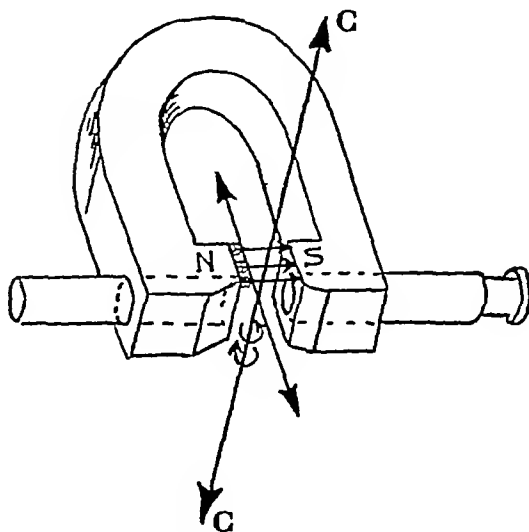


FIG. 23.1 C, C, galvanometer string, N and S, north and south poles of electromagnet. Further description in text.

(or foot) These may be coupled in any one of three combinations, each of which is referred to as a *lead*. They are—

Right arm and left arm = *lead I*

Right arm and left leg = *lead II*

Left arm and left leg = *lead III*

These are the leads originally employed by Einthoven, and the only ones in use clinically for many years, they are known as the *standard leads*, or as *bipolar limb leads*, and because they record differences in potential between two points (two arms or an arm and a leg) remote from the heart they are also referred to as *indirect leads*.

The electrodes consist of pliable metal electrodes (3" x 2" in area) covered with gauze soaked in strong salt solution or with a jelly composed of salt, tragacanth, glycerine and water. They are applied to the forearms and the left leg. By means of a switch on the control board of the electrocardiograph, a record from each lead can be taken in turn.

Unipolar direct and semidirect leads Wilson in his experiments on bundle branch block in the dog's heart (ch. 24) observed that when one electrode was placed on the chest wall overlying the heart, and the other situated on some remote indifferent part of the body, the electrocardiogram was almost identical with one obtained when the first or exploring electrode was in contact with the cardiac surface. The potential changes occurring beneath the *distant* or *indifferent* electrode are of small magnitude, and overshadowed by the large potential changes occurring beneath the precordial or exploring electrode. The indifferent electrode should have as large an area of contact as possible, the potential changes being then minimized, a large area of the right arm is usually selected.¹ Whereas in the use of bipolar indirect leads the potential difference between two remote parts is recorded, when a unipolar lead is employed the potential changes of a relatively small area of the heart beneath the exploring electrode is recorded.

In addition to the standard leads one or more records are now usually taken by the unipolar method, the exploring electrode being placed on the precordium or a neighboring area of the chest. Such leads in close proximity to the heart are called *semidirect*, if the exploring electrode were placed directly on the surface of the heart (which in man is possible, of course, only under rare and special circumstances) the lead would be

¹ In Wilson's method of employing unipolar semi-direct leads, the exploratory electrode is paired with a central electrode which is formed by connecting together all three standard leads—both arms and left leg—through 5000 Ohm resistances. This serves as an "artificial ground" or neutral electrode, the potentials of the three leads tending to cancel one another (fig. 23.3)

called *direct*. Records taken with semidirect leads give more information concerning the heart, especially with respect to small myocardial infarcts, ventricular hypertrophy or bundle branch block, than do the standard leads. These unipolar chest leads are designated by the letter C, the position of the indifferent electrode by R, L, F or B (to indicate, respectively, right or left arm, left leg or back), and subscript numbers 1 to 6 for the several positions of the exploring electrode in most common use. If Wilson's central electrode is employed (see footnote) the unipolar chest leads are designated V with subscript numbers 1 to 6. These designations are listed below and are shown in figure 23.4, p. 220.

CR₁, CL₁, CF₁, CB₁, (V₁), right 4th intercostal space at the border of the sternum

CR₂, CL₂, CF₂, CB₂, (V₂), left 4th intercostal space at the border of the sternum.

CR₃, CL₃, CF₃, CB₃, (V₃), midway between V₂ and V₄

CR₄, CL₄, CF₄, CB₄, (V₄), over apex beat or left midclavicular line in 5th intercostal space

CR₅, CL₅, CF₅, CB₅, (V₅), midway between V₄ and V₆

CR₆, CL₆, CF₆, CB₆, (V₆), in left midaxillary line at level of apex beat.

Continuing around the chest at the same level, three other positions of the exploring electrode, namely, in the left postaxillary line, at the angle of the left scapula and over the vertebral column, they are designated, respectively, by subscribing the numbers 7, 8 or 9 to the letters CR, CL or CF or to V. In other instances, the exploring electrode is inserted into the esophagus where it lies in close proximity to the heart. This esophageal lead is lettered *VE*, a subscript number indicates its distance, in centimeters, from the teeth (e.g., VE₁₅).

Unipolar limb leads A unipolar limb lead is one in which the exploring electrode is upon an arm or a leg, and paired with a central electrode as described in the foot note on this page, or the central electrode may be made by connecting two instead of three limb leads—the so-called augmented unipolar limb lead of Goldberger—the connection with the limb lead to which the exploring electrode is connected being omitted. The deflexions so-obtained are of larger amplitude than when the three leads are joined. Unipolar limb leads are designated V_R, V_L, and V_F for right arm, left arm and left leg positions, respectively, of the exploring electrode.

A sagittal bipolar chest lead—in which one electrode is placed on the 4th intercostal space to the left of the sternum, the other on the posterior wall of the chest directly opposite the first—is recommended by Wolferth and Wood. They have shown that abnormalities may be detected in this plane which do not appear in records taken with the other leads.

An endocardiac (direct) lead can be employed but its use is confined chiefly to clinical research. The direct lead is introduced into the auricle or ventricle by catheterization through the median basilic vein (p 268). The electrogram so obtained differs from the ordinary electrocardiogram. The auricular complex is represented by a series of 4 rapid deflections (P wave) above and below the base line, followed by a negative T_a wave. The ventricular complex as recorded from the right ventricle consists of a positive deflection R, a negative S, followed by a second positive wave R'. The T wave may be negative or positive (see Battro and Bedoggia)

THE NORMAL ELECTROCARDIOGRAM (ECG)

During the cardiac cycle the electrical changes in the heart cause five distinct movements or deflections of the galvanometer string. Three of these are directed inward, i.e., toward the arch of the magnet, two are directed outward. The photographic surface, as we have seen, moves perpendicularly from above downward. The movements are, in consequence, marked in the record as deflections to one or other side of a vertical zero line representing the isoelectric position of the string. The electrocardiogram, however, is always examined after rotating the plate to the left into the horizontal position, when the deflections appear as waves above or below the zero line. The waves from left to right have been designated simply by the letters of the alphabet, P, Q, R, S and T (fig 23 5). P, R and T are directed upward, Q and S downward. Q is not always present. A fourth small positive wave, U, is sometimes seen in leads I and II immediately following T. Its significance is unknown. The vertical lines in this figure mark fifths of seconds, the horizontal lines, 1 mm apart, indicate tenths of millivolts.

The following is a description of the waves taken with standard leads

The P wave is produced by the spread of the excitation wave over the auricles. Notching of the P wave not infrequently occurs in leads II and III, and occasionally in lead I. The duration of the P deflection as measured along the base line averages 0.09 second, and normally does not exceed 0.12 second. It precedes the auricular contraction by about 0.02 second. A negative wave following P is sometimes seen, though it is usually buried in the QRS complex. It is called the T_a wave (T wave of auricle). It is more likely to be evident in heart block, when an auricular contraction occurs without being succeeded by ventricular systole.

The Q, R, S and T deflections are produced by the

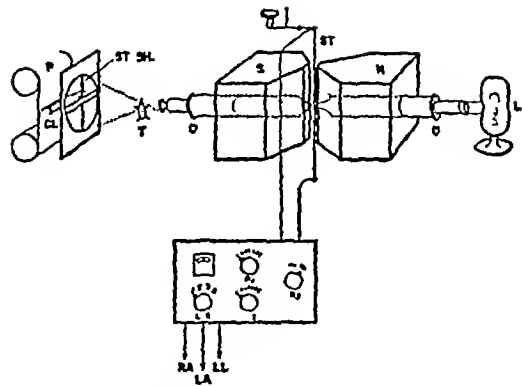


FIG 23 2 Diagram of the principal parts of the electrocardiograph. P, roll of photographic paper moving behind the lens, CL, which narrows the shadow, ST, SH, of the string ST to a thin vertical line and focuses it upon the sensitive paper. By means of ruled etchings upon the lens, horizontal lines 1 and $\frac{1}{2}$ mm apart are thrown upon the record, the heights of the waves in the electrocardiogram can thus be measured and the electromotive force in each wave calculated. T is a rotating toothed disc which breaks the light beam at regular intervals of $\frac{1}{50}$ or $\frac{1}{25}$ seconds and by throwing vertical shadow lines upon the record serves as a time marker. O and C indicate the positions of the lenses in the projection system, S and N are the poles of the electromagnet, L is the source of light. The current from the subject passes through the control box below. Switches on the dial enable a record from any lead to be taken. RA, LA and LL represent the three standard leads, right arm, left arm and left leg (Kindness of Dr L N Katz)

ventricles, the first three waves during the spread, and the T wave during the retreat of the excitation wave (p 201). The duration of the QRS complex varies inversely with the heart rate. In normal young adults it ranges between 0.05 and 0.1 at the ordinary heart rate. A duration exceeding a tenth of a second is abnormal.

The R wave commences about 0.02 second before the beginning of the rise in the intraventricular pressure curve (p 207), it reaches the base line again before the semilunar valves open.

The duration of the P-R interval (from the beginning of P to the beginning of QRS) varies inversely with the heart rate. The normal range at ordinary heart rates in normal young adults is from 0.12 to about 0.20 second. It increases in length with age, ranging between 0.07 and 0.14 at 1 year and between 0.14 and 0.23 second in persons over 60. The length of the interval is taken as a measure of the time taken for the impulse to pass over the auricle and A-V connections as far as the upper part of the interventricular septum. A value exceeding the upper normal limit indicates an abnormally slow conduction rate (p 227).

The S-T interval, that is, the interval elapsing between the end of the S wave (or of the R wave when the S wave is absent) and the commencement of T during which the record follows the base line, varies

in length from zero to 0.15 second according to the rate of the heart. It coincides approximately with the period of maximum ejection of ventricular systole (p 208), when the entire ventricular musculature is contracting and therefore in a state of equal electrical potential. The T wave represents the time of reduced ejection, when the electrical activity of the heart is subsiding. Its termination coincides with the commencement of isometric relaxation. A measurement of the Q-T interval (beginning of Q to end of T) gives very approximately the length of ventricular systole. The Q-E interval, i.e., the lapse of time between the commencement of the Q wave and the commencement of ventricular ejection as determined from the subclavian pulse, is from 0.10 to 0.16 second.

The U wave which follows the T wave by about 0.03 second, is a low broad deflection occurring in 60 per cent or more of normal electrocardiograms. It has a duration of about 0.2 second, and is thought to be due to the supernormal phase of excitability of the ventricular muscle, and to correspond, therefore, to the after potential of nerve. Most ventricular extra systoles occur at this time. The U wave is likely to be more pronounced in athletes (i.e., to have a higher voltage), or when, from any cause, the stroke volume of the heart is increased.

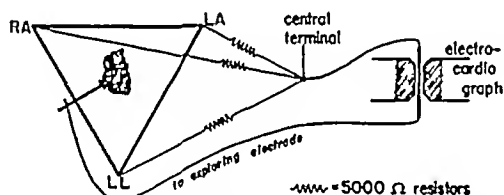


FIG 23.3 Showing central electrode (Wilson) After Hecht, *The Basic Principles of Clinical Cardiology*, by permission of C. C. Thomas, publisher

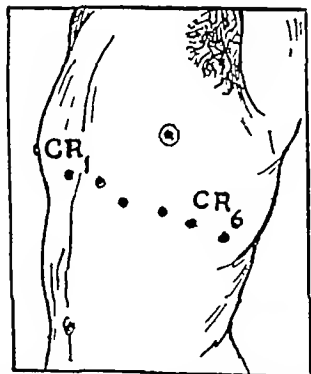


FIG 23.4 Showing positions of chest leads. Leads between I and 6 in numerical order from left to right

Voltages The highest wave of the electrocardiogram (R wave) in lead II ranges in normal young adults at rest and recumbent from 0.5 to 18 mm (i.e., from 0.005 to 1.8 millivolts (mV)). The heights of the other deflections of the electrocardiogram (standard lead II) of young adults are given below

P wave	0.2-2.2 (mean 1.0)✓
Q wave	0.06-1.9 (mean 0.3)
S wave	0.00-4.9 (mean 1.2)
T wave	0.4-5.6 (mean 2.9)
U wave	0.013-0.05 (mean 0.03)

THEORY OF ELECTROCARDIOGRAPHIC INTERPRETATION

EINTHOVEN'S TRIANGLE

Let a plate of conducting material be fashioned in the form of an equilateral triangle (fig 23.6, I) and let any two apices of the triangle be connected through a galvanometer so that when the fall in potential is from A to C either through B or directly to C the galvanometer deflection is upward, i.e., positive, and when the fall in potential is in the opposite direction the deflection is downward or negative. When an electro-

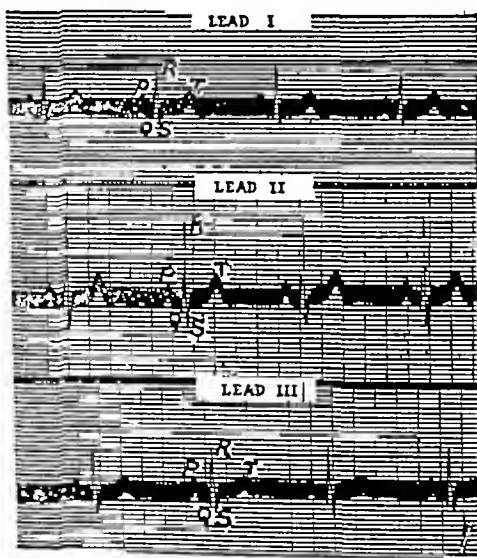


FIG 23.5 Normal electrocardiogram (kindness of Dr John Hepburn)

motive force, whose magnitude and direction (vector) are represented by the length and direction of one or other of the central arrows shown in figure 23 6 I, is developed within the triangle, then the potential difference between any two of the latter's apices will be proportional to the length of the projection of the arrow upon the side connecting these apices. The length of the projection on a given side of the triangle, and so the recorded potential change which it represents, will vary with the angle which the arrow makes with the horizontal. Whether the central EMF produces an upward or a downward deflection will depend upon its general direction. In the figure all the projections of the solid arrow are positive, two of the projections of the arrow drawn in interrupted lines are negative and one positive. But the arrow might point in any direction, and depending upon its direction in relation to a given side of the triangle its projection upon that side will be of positive or negative sign. It will be found, nevertheless, that in whatever direction the arrow points, the algebraic sum of its projections on the sides AB and BC will equal its projection on AC.

If the source of the electromotive force were in the center of a conductor of three dimensions (volume conductor), e.g., a body of fluid containing electrolytes, vectors in any radius could be calculated.

Now let the three standard electrocardiographic leads be imagined as forming an equilateral triangle about the heart (fig 23 6, II), let the electrical axis or vector² of the heart be represented by the arrow, and the potential differences, negative or positive, which are recorded (as deflections) in the electrocardiogram in the three leads, be taken to correspond to the projections of the arrow upon the respective sides of the triangle. Then the potential difference between any two points upon the body represented by the apices of the triangle, and recorded in any lead and at any instant, will vary in magnitude in accordance with the angle (designated angle α) which the arrow makes with the horizontal, and will be proportional to the size of the projection of the arrow upon the line of the lead. Then the algebraic sum of the potential differences as recorded in leads I and III will equal that recorded in lead II. This truth, first enunciated by Einthoven (Einthoven's law), is expressed by the following formula, $e^1 + e^3 = e^2$, in which e represents potential difference and the numerals the respective leads. It forms the basis for the interpretation of the electrocardiogram.

It may be stated further and in general terms that if three points (whether or not they form an equilateral triangle) are equidistant from a source of energy (e.g.,

a flame or the heart generating electric currents) the difference between the values at points 1 and 3 can be predicted if the differences between points 1 and 2, and 2 and 3 are known. Therefore, knowing the voltages of a deflection in lead I and lead III (potential difference between right arm and left arm, and between left arm and left leg, respectively) the voltage in lead II can be calculated.

THEORIES OF THE ORIGIN AND SPREAD OF THE ELECTRICAL CURRENTS IN THE HEART

From the time that the electrical currents of the heart were first demonstrated by Koelliker and Muller (1856), and the discovery by A. D. Waller (1879) that they could be recorded in man, several theories have been proposed as to the mode of their production and to explain how they are transmitted and recorded as the characteristic deflections of the electrocardiogram.

Starting with the discharge of the impulse and its transmission over the auricular muscle, leaving unexplained the underlying chemico-physical changes leading to its generation within the pacemaker of the heart (SA node in mammals), a theory known as the base-apex theory, or theory of distributed potential differences, was offered. This was founded upon what was then known of the principles of electro-physiology and the anatomy of the heart. It was supposed that the wave of excitation, having traversed the auricular muscle and produced the P wave, and then entering the ventricular muscle, followed a course from the base of the right ventricle to the apex and then from the apex to the left ventricular base. It was believed by Gotch that such a pathway was determined by the developmental history of the heart, namely, the doubling upon itself of the embryonic cardiac tube. The region at the base of the right ventricle was the first to become electrically negative³ to the rest of the heart and produced the R deflection. The apex was next excited and the potential change re-

³ To avoid confusion, the term electrically negative should be defined. Injured or active tissue by convention is said to be electrically negative to uninjured and resting tissue. But the active or injured tissue is comparable to the negative pole of a voltaic cell of which the two poles are, let us say, of zinc and copper. But zinc, though the negative pole, is actually electro-positive, the current flowing within the battery from zinc to copper and from copper to zinc outside. Lewis suggests the use of the word "zincative" when referring to active or injured tissue.

² A vector is anything, such as force, velocity, momentum or electrical energy having both direction and magnitude and which can be represented by a straight line of appropriate direction and length.

versed, the S deflection was then inscribed, i.e., it became electrically negative to the base and the rest of the heart. Activity next spread to the base of the left ventricle and persisted here for a time after it had ceased at the apex. Thus the T wave was formed. In this way, the spread of the excitation wave in the heart was compared to the spread in a stretch of skeletal muscle.

But subsequent research proved the inadequacy of this simple explanation which was based largely upon experiments with the amphibian heart in which electrodes were placed directly upon base and apex.

The excitation wave does not spread in this way in the mammalian heart, but is transmitted rapidly through a special conducting system, whose terminal ramifications penetrate to all parts of the cardiac musculature. It does not follow the muscle bundles (p. 201) but reaches innumerable points in the relatively large areas of the myocardium almost simultaneously. Furthermore, even when electrodes are placed upon the base and the apex of the frog's heart, the deflection does not always indicate that the muscle beneath the basal electrode is activated first. It depends upon the precise location of the basal electrode, some points at the base are activated before, others after activation of the apex.

According to the theory of *limited potential differences* proposed by Lewis, minute areas of muscle become electrically negative to corresponding areas lying immediately adjacent. The rest of the myocardium is considered to be electrically neutral, so microscopical electric cells, called *dipoles* or *electrical doublets* by Craib, are visualized, which constitute the advancing front of the wave of excitation.

In contrast to the theory of *distributed potential differences*, in which the differences of electrical potential are created between relatively large masses of the myocardium—base and apex—the theory of limited potential differences considers the heart muscle as consisting of innumerable units in which action currents having different directions at any given instant and from one instant to the next, are developed. Some of these will neutralize, others will reinforce, one another. It is the algebraic sum of these potential changes which determines the potential differences as recorded by the electrocardiograph. It can be represented by a line—a *vector*—having a certain definite direction and magnitude. The resultant of the electromotive forces developed within the heart and recorded at any instant is spoken of as the *electrical axis of the*

heart. It is the direction and magnitude of current flow, i.e., of the electrical axis rather than the location relative to one another of two oppositely charged *masses* of cardiac muscle, which determines the characters of the electrocardiogram.

A general theory of the genesis of the electrocardiogram founded upon the classical *membrane theory* (chap. 62) seems to be in best agreement with the physiological and anatomical characteristics of cardiac muscle. The membrane theory in its general application is as follows. The electrolytes within any living cell differ in kind and in concentration from those in the surrounding fluid. The intracellular and extracellular constituents are separated from one another by a semipermeable membrane—permeable to some ions but relatively, or absolutely, impermeable to others. Ions became aligned on the two sides of the membrane—those on the inner side being negatively charged, those on the outer side carrying a positive charge. Thus an electrical potential is established between the two surfaces of the membrane. The polarity of the membrane is shown in figure 23.7. When a stimulus is applied to the cell, membrane permeability is reduced, a leakage of ions across the membrane follows, i.e., depolarization occurs, and a current flows within the cell from the depolarized region to the adjacent part which is still in the polarized state, and in the opposite direction on the other side of the membrane (fig. 23.7). Activity, then, is associated with depolarization, recovery with repolarization. (The mechanical change in the cell follows upon the electrical change and is not responsible for it.) No electrical changes are created in the resting, uninjured state of the cell, when its membrane surface is completely polarized or when it is completely depolarized. An electromotive force is produced only during depolarization or repolarization.

Now the cardiac muscle is a syncytium, that is, a multinucleated protoplasmic mass, being undivided like skeletal muscle into small units by membranous partitions. The auricles taken together may, therefore, be considered as a single cell and the two ventricles another. These two cell units are connected only by the junctional tissues. Consequently, changes in electrical potential will occur only when depolarization occurs at some part of the cardiac surface—endocardial or epicardial—where the myocardium is covered by membrane. Only when the impulse "breaks through" to some part of one or other surface will an electrical change be recorded in the elec-

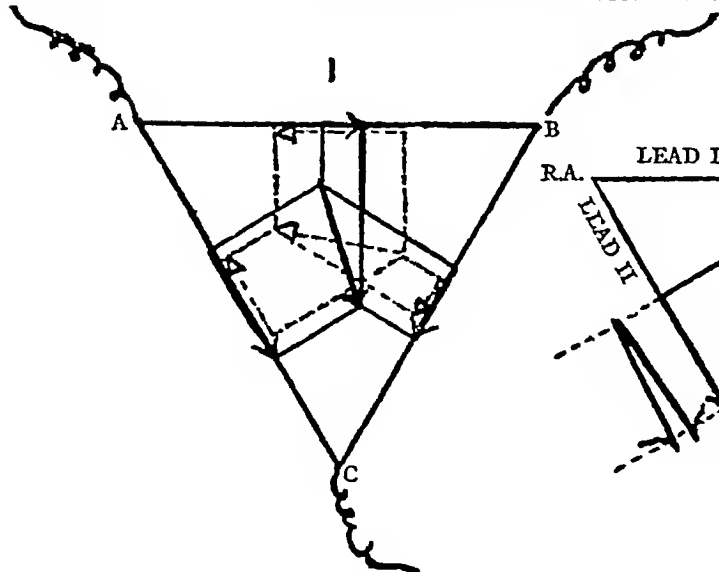


FIG 23 6 Description in text.

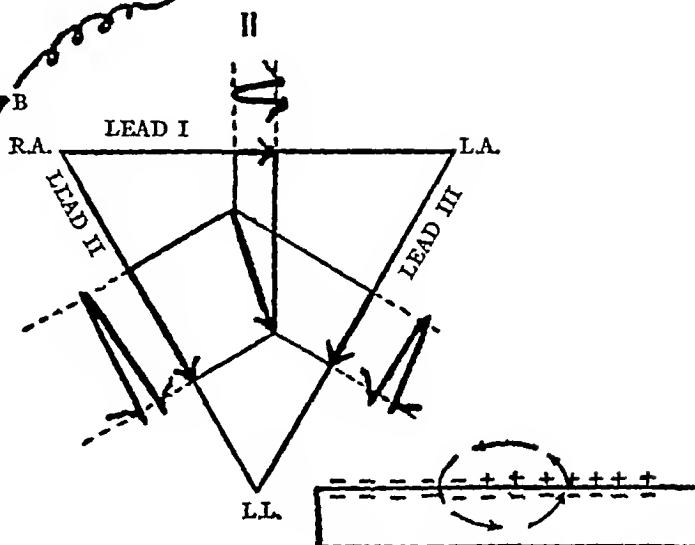


FIG 23 7 Description in text

trocardiogram. Thus, an explanation is given for the fact that, according to Katz, an injury within the substance of the myocardium gives no electrocardiographic evidence of its presence unless it extends to the endocardium or epicardium. The course of the processes of depolarization and repolarization in time and space are reflected in the contours of the electrocardiogram. Though they cannot be followed precisely, a certain degree of orientation of the course of the electrical events has been shown. Depolarization spreads from the S-A node through the auricular muscle, causing the P wave, repolarization commences almost immediately after and follows approximately the same course as depolarization, but progresses more slowly. It gives rise to the T_n wave. Reaching the upper part of the septum on either side beneath the endocardium, the impulse causes depolarization. The Q wave is inscribed. It then spreads through the septum and, by the arborizations of Purkinje, to all parts of the ventricular muscle, in which large areas are activated at innumerable points almost simultaneously. Depolarization spreads to the apex, then to the muscle in the interventricular grooves, then to the lateral walls of the right and left ventricles, and finally to the conus region of the right ventricle. As the depolarization process takes this course, the R and S deflections are produced. The depolarized state persists for an appreciable time, and, since it involves the entire surface of the ventricles, no potential difference exists and the electrical record follows the base-line (isoelectric or isopotential line). This constitutes the S-T interval or segment. Repolarization then commences, and proceeds relatively slowly to create an electromotive force which gives rise to the T wave.

The changes in electrical potential developed at the surface of the heart set up an electric field in the surrounding conducting medium (tissue fluids and blood), which spreads eccentrically in a direction at right angles to the course taken by the depolarization or repolarization process.

THE ELECTRICAL AXIS OF THE HEART

The electrical axis of the heart has been defined on page 223 as the resultant of the electromotive forces developed in the heart and recorded at any instant.

Determination of the electrical axis of the heart, vectorcardiogram, spatial electrocardiogram

In the method of Einthoven and Fabr the electrical axis of the heart projected in the frontal plane of the body is determined as follows. A deflection or group of deflections (e.g., QRS) in each of the three standard leads is drawn to scale on the corresponding side of the triangle. Perpendiculars are dropped from simultaneous points on the deflections toward the center of the triangle. A line joining the intersections of these perpendiculars gives the direction, and its length the magnitude of the potential difference between the simultaneous points as projected in the frontal plane of the body. Any number of such vectors can be drawn from a corresponding number of simultaneous points. Three are shown in figure 23 6, each is known as an instantaneous electrical axis. If the distal ends of the lines are joined a loop is formed the long axis of which is called the modal electrical axis, it indicates the maximum manifest potential value of the electrocardiographic deflection whose axes are represented. If a large number of instantaneous axes are drawn and their ends joined the loop so formed is irregularly elliptical in shape, and is called a vectorcardiogram in the frontal plane. If instantaneous electrical axes are determined in a similar manner in the sagittal plane, and the

vectorcardiogram combined with one in the frontal plane, a three dimensional or *spatial* vectorcardiogram can be constructed

The rotation of the electrical axis

Determinations of the electrical axis at successive moments throughout the cardiac cycle have shown that its direction is continually changing. This follows naturally from the fact that the electrical axis is an expression of the balance that has been struck at the moment between separate action currents. Since the directions of these vary from instant to instant, the direction of the resultant force would be expected to vary likewise (fig 23.8). During the spread of the excitation wave in the heart the electrical axis swings in an orderly manner from left to right.

DEVIATION OF THE ELECTRICAL AXIS

The direction of the electrical axis, of course, bears a relationship to the anatomical axis of the heart. The latter may be taken as passing longitudinally through the interventricular septum, being directed forward, downward and to the left and roughly parallel to the side of the triangle represented by lead II. Obviously, changes in the position of the heart in relation to the sides of the triangle enclosing it will, by causing corresponding alterations in the direction of the electrical axis, be reflected in the electrocardiogram. The heart alters its position slightly during ordinary respiration. During expiration, as a result of the ascent of the diaphragm, the apex swings upward and to the left, i.e., anti-clockwise, and the heart assumes a more transverse position, the opposite movement occurs during inspiration.

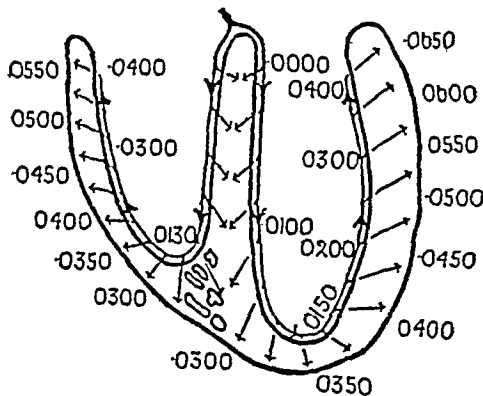


FIG 23.8 A diagram of the human heart in section, representing the directions in which the excitation wave spreads in the human ventricle and the time in seconds at which, after its commencement in the ventricle, the wave first reaches various regions of the ventricle. (After Lewis.)

Even changes in the position of the body (turning in bed from one side to the other) or distension of the stomach may cause slight changes in the direction of the anatomical axis. A rotation of the heart to the left tends to alter the direction of the electrical axis (*left axis deviation*) in such a way as to increase the amplitude of the R wave and reduce the S wave in lead I, and to reduce the R wave and increase the S wave in lead III. The more transverse position of the heart tends also to reduce the P and T waves in lead III or to cause inversion of the T wave. Movement of the heart into a more vertical position produces the opposite effects (*right axis deviation*) upon the QRS complex—reduction in the height of the R wave in lead I with an increased depth of the S wave, a small P and an upright T. In lead III the amplitude of R is increased. Rotation of the heart may result from tumors, hypertrophy of the ventricle, etc. In the rare developmental anomaly—complete transposition of the heart (*dextrocardia*)—inversion in lead I of all the waves of the electrocardiogram occurs. A similar ECG is produced if the arm leads of a normal person are reversed, one taken from the right precordium is the mirror image of the normal taken from the left side. The heart lies with its apex pointing to the right. Displacement of the heart as a whole, as by a pleural effusion or pneumothorax, is less likely to produce changes in the electrocardiogram than when it undergoes rotation.

It has been mentioned above that division of one branch of the bundle disturbs the electrical balance of the heart, an abnormal deviation of the electrical axis to the right or left will be a result, interruption of a branch of the bundle by disease causes a similar distortion of the electrocardiogram (see p 229). In cardiac hypertrophy and ventricular extrasystoles abnormal deviation of the axis will also occur. Left bundle branch block, left ventricular hypertrophy and right ventricular extrasystoles (see fig 23.9, p 225) cause left axis deviation, right bundle branch block, right ventricular hypertrophy and left ventricular extrasystoles (see fig 23.10) produce right axis deviation.

There has been a considerable amount of discussion concerning the mode of production of the electrocardiographic features of ventricular hypertrophy. The following explanations have been advanced

- (1) That it is due to the greater mass of muscle,

the electrical changes occurring in the hypertrophied chamber overbalancing those of the normal side. This theory has been disproved by the discovery that the electrocardiographic features which were thought to be produced by one ventricle are actually produced by the other (p 229). In other words, the electrical changes in the sound ventricle overbalance those of the hypertrophied side.

(2) Altered position of the heart The greater mass of the hypertrophied ventricle causing rotation of the heart around its longitudinal axis, this probably plays only a minor rôle.

(3) Slowed conduction over the bundle branch supplying the hypertrophied ventricle as a result simply of the lengthening of the conduction pathway incident to the enlargement of the ventricular cavity, or to actual injury of the conducting tissue associated in some way with the cardiac disease. According to Barker and associates one or other of these two factors is the most probable explanation of the characters of the electrocardiogram in ventricular hypertrophy.

Conditions associated with flattening or inversion of the T wave Vagal and sympathetic stimulation may be followed by inversion of the T wave. Inversion also appears in the early stages of digitalis poisoning, after the administration of adrenaline, quinidine and other drugs, and in acute infections. In these instances there may be little or no alteration in the QRS complex. Inversion of the T wave in leads I and II is taken in general as a very unfavorable sign and frequently accompanies grave cardiac disease. But though it is true that in a series of cases showing persistently an inverted T the duration of life is shorter than in those in

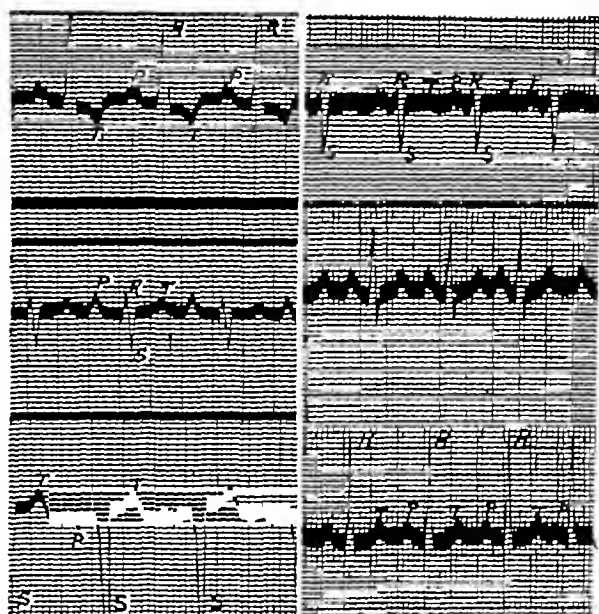


FIG 23.9 Abnormal axis deviation a, left ventricular hypertrophy, b, right ventricular hypertrophy, vertical lines mark $\frac{1}{2}$ sec. (kindness of Dr John Hepburn)

which it is upright, the direction of the wave nevertheless cannot be relied upon alone in arriving at the prognosis of an individual case. The deflection may be inverted in temporary and comparatively unimportant conditions, or it may be upright though a fatal termination from cardiac failure is imminent. Flattening or slight inversion of T in lead III is of no significance, it may occur in perfectly normal persons.

THE VENOUS PULSE

Pulsation in the large veins at the root of the neck is a normal phenomenon which can be demonstrated by means of special apparatus. It is only under abnormal conditions, however, and then

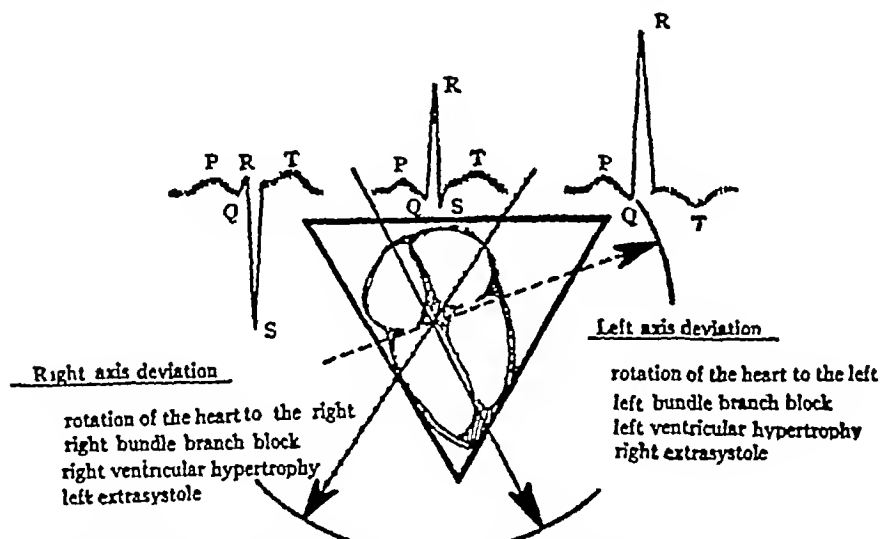


FIG 23.10 Diagram to illustrate right (beaded arrow) and left (interrupted arrow) axis deviation. Normal electrical axis shown by arrow with solid shaft. Electrocardiograms are for standard lead I.

only occasionally that the pulsations are visible to the eye. Valuable information concerning the events of the cardiac cycle from an experimental as well as from a clinical point of view can be gained by a study of jugular pulse records.

The pulsations occurring in the jugular bulb are the result of pressure changes occurring within the right auricle. Since the vein is in direct communication with the interior of the auricle, changes in volume and pressure, which correspond closely to the pressure fluctuations of the intra auricular pressure curve, are transmitted to the column of venous blood. The intra auricular pressure is influenced by events occurring not only in the auricle itself but by pressure changes transmitted from the ventricle. For this reason a study of the waves of a jugular tracing enables the true relations of the different events of the cardiac cycle (auricular and ventricular) to be determined. The jugular tracing, though the counterpart in a qualitative sense of the intra auricular pressure curve, gives no quantitative information, being a record of volume and relative pressure changes only, and no criterion of the absolute pressure values developed in the auricle. The waves c and v represent ventricular events and are in consequence spoken of as the ventricular complex. The interval between the commencements of the a and c waves (a-c interval) indicates the time elapsing between auricular and ventricular systoles and its duration is an index of the conduction rate between the two chambers.

The waves of the venous pulse are given the same lettering as those of the auricular pressure curve (p 205). Their main features are

Positive wave a due to auricular systole.

Negative wave x due to commencement of auricular relaxation.

Positive wave c due to ventricular contraction and bulging of A-V valves into the auricle.

Negative wave x' due to drawing down of A-V septum and the discharge of blood from the thoracic cavity.

Positive wave v due to filling of auricle which is closed below by A-V valves.

Negative wave y due to opening of A-V valves and emptying of auricle into the ventricle.

In figure 23 11 is shown a typical venous pulse. The waves show their distinctive characters which may be recognized almost at a glance. Venous tracings from a clinical case, however, are frequently atypical and present a confused series of irregular waves which are impossible to identify by mere inspection. Sometimes successive waves, for instance, a and c, or v and a, are

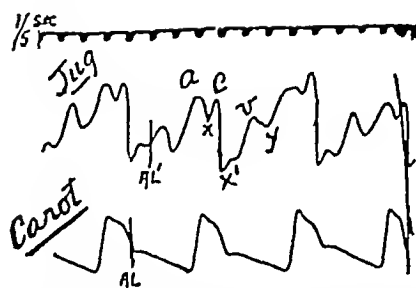


FIG 23 11 Upper tracing from jugular vein, lower from carotid artery AL, alignment marks (after Price)

merged together. At other times certain waves are absent. We must then have recourse to some method of distinguishing the various fluctuations in order to interpret the tracing. An arterial tracing serves as a key. In figure 21 3 (p 207) the intracardiac, venous, and arterial curves are accurately superimposed, that is, they all commence at the same instant. Consequently synchronous events lie along vertical lines intersecting the several tracings. In the case of the jugular tracing the commencement of the upstroke of the carotid is synchronous with the commencement of the c wave. The upstroke of the carotid or of the radial artery is practically always a landmark which is clearly discernible. Therefore, if one could take a venous tracing accurately superimposed with a carotid tracing a vertical line drawn through the commencement of the upstroke of the latter would when extended through the venous curve indicate the commencement of the c wave. It is not feasible, however, to do this, but the time relations of the two writing levers can be correlated by means of what are known as alignment marks. That is, while the writing surface is at rest the two levers are given a light tap so that each makes an upright mark. The marks are used as points from which measurements may be made and the two tracings synchronized in an indirect way. Thus (cf fig 23 11) the distance from the alignment mark AL to the carotid upstroke is measured. This distance is then laid off upon the venous tracing commencing from the alignment mark AL. It will indicate a point on the jugular tracing corresponding to the commencement of the c wave. When the arterial tracing is taken from the radial, as is most commonly done, the procedure is the same, except that a distance representing $\frac{1}{6}$ second, obtained from the time tracing, must be deducted from the measurement. This is the difference in the times of arrival of the pulse at the carotid and the radial. The instrument employed for the clinical registration of jugular and arterial pulses, and originally devised by Mackenzie, is known as the polygraph.

In determining the length of the arterial pulse wave (p 184) a procedure similar to that just described is followed in timing the successive arrivals of the wave at two points in the arterial tree.

CHAPTER 24

DISORDERS OF THE HEART BEAT AND THEIR INVESTIGATION BY GRAPHIC METHODS

The following is a convenient classification of cardiac irregularities

A *Affections of rhythm due to impaired conduction through the A-V bundle and its ramifications*

- I Stem of bundle
 - (1) Delayed conduction
 - (2) Missed beats, partial heart block
 - (3) Complete heart block
- II Bundle branch block Intraventricular block

B *Wolff-Parkinson-White syndrome*

C *Affections due to abnormal impulse formation*

- I Extrasystoles
 - (1) Ventricular
 - (2) Nodal
 - (3) Auricular
- II Paroxysmal tachycardia
 - Auricular, nodal and ventricular
- III Auricular flutter
- IV Auricular fibrillation
- V Ventricular fibrillation

D *Alternation of the heart*

E *Affections due to vagal influences*

- I Sinus arrhythmia
- II Phasic irregularity
- III Sinus bradycardia
- IV Sino-auricular block
- V Auriculoventricular block

A *AFFECTIONS OF RHYTHM RESULTING FROM IMPAIRED CONDUCTION*

I *IN THE A-V NODE OR STEM OF THE BUNDLE—AURICULO-VENTRICULAR BLOCK*

In animals, conduction from auricle to ventricle can be depressed or blocked by crushing, cutting, or the application of cold to the A-V bundle. This strategic point in the pathway of the excitation wave is also attacked by disease, and conduction through it may be depressed or completely abolished. Depression of conduction through the node or bundle varies in degree. Three stages are recognized

(1) *Delayed conduction*

In this stage conduction is merely slowed, every impulse reaches the ventricle. The intervals be-

tween the auricular and ventricular systoles (A_s-V_s intervals) are lengthened and may have a duration of 0.5 second, though, as a rule, they are considerably shorter than this. The condition can be recognized only by means of the electrocardiograph or a venous pulse tracing. Lengthening of the P-R interval in the former tracing, or of the $a-c$ interval in the latter, beyond the normal maximum of 0.2 second is taken to indicate delayed conduction (fig. 24.1)

(2) *Missed beats—partial heart block*

When impaired conduction reaches a certain degree, impulses from time to time fail to reach the ventricle, and a beat is missed. The auricular beats are perfectly regular, and in this way the condition differs from sino-auricular block (p. 240). A ventricular beat may be missed only occasionally and at irregular intervals. The P-R interval of the electrocardiogram, or the $a-c$ interval of the venous pulse, may generally, though not invariably, be seen to lengthen progressively for several heart cycles preceding the dropped beat. The periods of delayed conduction preceding the missed beat are called, after their discoverer, the Wenckebach periods. The interval of the cycle succeeding the missed beat is shortened again to near the normal length. In a further stage of the condition the beats are dropped more frequently and may be spaced at either regular or irregular intervals in the tracing. When the grade of block is still more advanced, impulses fail to penetrate the bundle after every second auricular beat, or three, or even four auricular contractions may occur before an impulse reaches the ventricle, i.e., the ventricle responds only to every third or fourth auricular beat. So, an auriculo-ventricular rhythm becomes established in which the two chambers beat in the ratio of 2:1, 3:1, or 4:1. The first of these is seen most frequently, the second is the least common (fig. 24.1)

(3) *Complete heart block*

When the A-V node or bundle offers an absolute barrier to the passage of the impulse, the dissociation of the rhythms of the two ventricles is com-

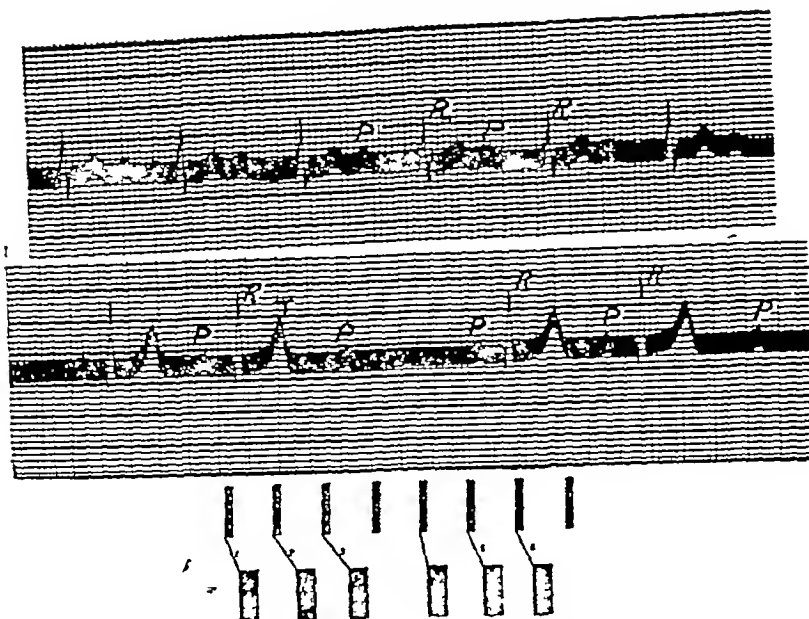


FIG 24.1 Upper tracing, delayed conduction, lead I Lower tracing, incomplete heart block, lead I (missed beats) (Kindness of Dr John Hepburn) The diagram below (after Lewis) represents incomplete heart block. The thin rectangles, A, represent contractions of the auricle the thicker ones, V, contractions of the ventricle. The obliquely directed lines represent conduction over the A-V bundle, the steeper the line the more oblique the line. The gaps in the lower rectangles indicate missed beats of the ventricle. It will be noted that delay in conduction increases progressively in successive cycles until a beat is missed. Heavy vertical lines = $\frac{1}{4}$ sec.

plete (fig 24.2) The auricles beat at their own rate of about 70 per minute and the ventricle at its inherent rate of about 35. The latter is then spoken of as the idio-ventricular rhythm. Both ventricles beat simultaneously. This fact suggests that when the ventricle assumes this rate it is under the control of some single region possessing the power of rhythmical activity. From experimental investigation it appears that the controlling center is the A-V bundle below the site of the lesion. The speed with which a particular region of the heart can develop and discharge impulses apparently determines its ability to dominate other regions. It has been mentioned that when a region of higher rhythmicity is destroyed or isolated the region next in order of rhythmical power assumes the rôle of pacemaker (pp 198, 242). When, for instance, the S-A node is destroyed or isolated the A-V node assumes control, and when this or the upper part of the bundle is separated from the tract of tissue below, the lower lying portion takes over the government of the ventricular rate.

Partial and complete heart block are accompanied by changes in the rhythm of the arterial pulse. When

beats are missed occasionally the pulse intermits either at regular or irregular intervals. In the more fully developed conditions marked slowing (bradycardia) of the pulse occurs. Visible pulsations in the veins at the root of the neck may occur, for the auricle in contracting upon a larger volume of blood accumulated as a result of the infrequency of the ventricular contractions, causes a pronounced wave to be transmitted along the jugular. A certain proportion of the venous pulsations may be seen to be unassociated with an arterial pulse. Sometimes a sound may be heard over the heart at the time of the isolated venous pulsation since the auricular contraction is unusually forceful and the sound vibrations thus set up are not smothered by the first heart sound which, in the normally beating heart, follows so closely upon auricular systole.

The venous pulse tracing and electrocardiogram show characteristic features. The a and P waves of the respective records occur at the usual times, but the ventricular complex (c and r in the venous curve and QRS and T in the electrocardiogram) is absent, a gap appearing in the tracing each time a beat of the ventricle is missed. In incomplete heart block, as one would expect, a relationship between a ventricular complex and a preceding auricular wave can always be made out, whereas in complete block there is no relationship. In the venous pulse, for example, the a and

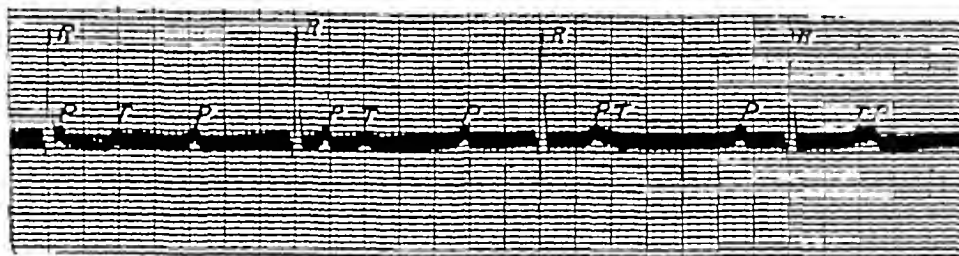


FIG 24 2 Complete heart block, lead I (kindness of Dr John Hepburn)

c waves may occur simultaneously, and produce a large $a + c$ wave. Or the a and v waves may coincide. Corresponding effects are produced upon the electrocardiogram.

Temporary heart block may result from various toxic agents, e.g., digitalis, strophanthus, quinidine, etc., which exert a specific effect in depressing auriculo-ventricular conduction. Heart block may be a sequel or an accompaniment of several acute infectious maladies, e.g., diphtheria, rheumatic fever, etc. It is produced in animals by asphyxia (p. 250). Increased vagal tone is sometimes responsible for delayed conduction over the A-V bundle. Partial heart block is not uncommonly seen in the course of rheumatic fever and is then, in many instances, of vagal origin, being temporarily abolished by atropine. *Persistent heart block* due to increased vagal tone and abolished by atropine occasionally occurs.

Stokes-Adams syndrome. This condition was first described by Adams (1827) and later by Stokes (1842). Its features are a slow pulse and syncopal attacks or convulsive seizures, usually epileptiform in character. The term includes any condition of vascular origin in which these features are associated, and it is probable that the underlying morbid state upon which the syndrome depends is not identical in all instances (see also carotid sinus, p. 283). In the majority, however, the slowed cardiac action is the result of heart block, and the cerebral symptoms are the direct result of the bradycardia. The prolonged pause between beats permits the diastolic pressure to fall to a low level, the blood supply to the cerebral centers suffers in consequence. Hardening of the larger arteries or aortic regurgitation, etc., when present in association with heart block, must enhance the effect of the bradycardia upon the diastolic pressure. That heart block alone is capable, however, of causing the syndrome was shown by Erlanger and Blackman (p. 201) who reported similar syncopal and convulsive seizures in dogs after division of the bundle. Adrenaline or ephedrin has been employed with some success.

II BUNDLE BRANCH BLOCK—INTRA-VENTRICULAR BLOCK

When one or other of the primary divisions of the bundle is blocked by disease the impulses

reach the muscle of the two ventricles asynchronously. The ventricle of the sound side is activated a fraction of a second before the other. The affected ventricle is excited later by the escape of the impulse from the healthy side through the septum (p. 222). The imbalance between the electrical effects in the two ventricles produces characteristic electrocardiographic features. Bundle branch block is most commonly associated with coronary disease. The left branch of the bundle is involved five times more frequently than the right (see also p. 224).

It may sometimes be of interest to know in which branch of the bundle the block occurs. Barker, MacLeod and Alexander stimulated the surface of the exposed human heart at different points and obtained extrasystoles (p. 230) whose characters indicated that a record in which the main initial deflection is upward in lead I and downward in lead III is due to disease of the *left* bundle branch and that an electrocardiogram in which the main initial deflection is downward in lead I and upward in lead III is due to disease of the *right* branch. Roberts and associates in experiments upon the cat and the monkey, also found that division of the right branch produced a record in which the main initial deflection was downward in lead I and upward in lead III. Opposite effects resulted from division of the left branch.

The direction of the T wave (upward or downward) is usually opposite to that of the main deflection of the QRS group. Thus in disease of the right branch the T wave has the normal direction in lead I but is inverted in lead III. Bundle branch disease is of the gravest significance, the subject rarely surviving many months after it has been discovered. The electrocardiograph is the only means by which it can be recognized. Distortion of the electrocardiogram similar in type to that seen in bundle branch block occurs in other conditions (p. 224) but a distinguishing feature of

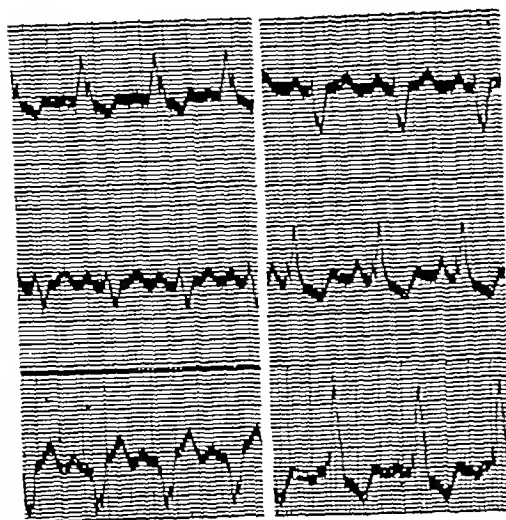


FIG 24.3 *a*, left bundle branch block, *b*, right bundle branch block. Note lengthening of QRS. Compare with fig 23.8 (kindness of Dr John Hepburn)

the former is lengthening of the QRS complex beyond the normal maximum of $\frac{1}{10}$ second. In disease of the left branch of the bundle the Q-E interval is prolonged owing to the lag in ventricular ejection.

B WOLFF-PARKINSON-WHITE (WPW) SYNDROME

In this rare congenital anomaly the P-R interval of the electrocardiogram is shortened and the QRS complex prolonged. The P-R interval is 0.12 second or less and the QRS complex extended to over 0.10 sec. The shortened P-R interval may be due, in some cases, to an accessory conducting strand between auricle and ventricle. Butterworth and Poindexter fashioned an extra electrical conduction pathway in the dog's heart, which, when stimulated, produced an electrocardiogram similar to that characteristic of the syndrome. Wood and Wolferth, in the histological examination of the heart of a patient which had shown this syndrome, also found three accessory pathways from the right auricle to the ventricle.

On the other hand, cases of the WPW syndrome occur in which no anatomical anomaly in the conducting system can be demonstrated, and even when such exists there is no conclusive evidence that it is the cause of the accelerated auriculo-ventricular conduction. Prinzmetal and his colleagues have produced typical WPW complexes in dogs by applying a non-interrupted direct current of subthreshold strength to the A-V node, but

such complexes could not be induced after severance of the A-V bundle—an indication that the impulses traverse the normal A-V connections. These investigators carried out a large number of similar experiments in which the movements of the heart were recorded by high speed cinematography, and the records analyzed after slowing down to half speed. They conclude from their observations that the accelerated conduction of the WPW syndrome is due to shortening of the delay of the impulse at the A-V node to a fraction of the normal which amounts to about 0.12 seconds.

These experiments have thrown considerable doubt upon the anatomical explanation of the WPW syndrome. Prinzmetal and his colleagues consider it to have a physiological rather than an anatomical basis.

The prolongation of the QRS complex is attributed to a functional bundle branch block.

C DISTURBANCES OF RHYTHM DUE TO ABNORMAL IMPULSE FORMATION

1 EXTRASYSTOLES OR PREMATURE CONTRACTIONS

An extrasystole can be induced experimentally by stimulating the cardiac muscle at any time except during its phase of absolute refractoriness (p. 189). Extrasystoles occur in the human heart as a result of some abnormal process of impulse formation. Though extrasystoles may be associated with organic heart disease, they more frequently occur in its absence; they may then be of reflex origin initiated from the abdominal viscera or be due to some form of intoxication, e.g., digitalis, chloroform anesthesia, hyperthyroidism, excessive smoking, etc. Beattie, Brown and Long produced extrasystoles in cats by stimulation of the hypothalamus (ch. 67), and their occurrence in man following brain lesions has been reported by Lucke and by Korth, which indicates that in some instances they are of central origin. The auricle or the ventricle may be the site of origin of the premature contraction, or the extra impulse may arise in the A-V node (see diagrams, fig. 24.4).

(1) Ventricular extrasystoles

The premature contraction occurs after the normal ventricular beat has ceased and the muscle has recovered from its absolute refractory state. It is not preceded by an auricular contraction, and is not dependent upon an impulse received from the upper chamber (fig. 24.5). The premature con-

traction is followed by a long pause. This is usually of just sufficient duration to cause the succeeding normal ventricular beat to occur at the instant that it would have occurred had there been no premature contraction. The cause of this *compensatory pause* has been explained elsewhere (p. 190). Briefly, it is due to the normal impulse reaching the ventricle when the muscle is still refractory as a result of the premature beat. The interval between the normal beat and the one following the premature contraction is therefore equal in length to two normal cardiac cycles (fig. 24.4, I). Sometimes, however, when the extrasystole occurs early in diastole and the heart rate is slow there may be no compensatory pause. The auricular impulse then reaches the ventricular muscle after it has recovered from the refractory state resulting from the premature beat, the auricular impulse therefore brings about a response at, or (as a result of some lengthening of the A-V interval following the premature beat) slightly after the usual time. The normal ventricular systoles are then all equally or nearly equally spaced and the extra contractions are interposed here and there between them. That is, the time interval from the normal beat preceding the extra contraction to that following it is of normal length or but slightly lengthened. Premature beats of this nature are called *interpolated extrasystoles*¹.

(2) Auricular extrasystoles

The premature contraction arises in the auricle at some point outside the S-A node. The abnormal impulse reaches the ventricle along the usual paths and evokes a ventricular contraction, unless the auricular beat is so premature that sufficient time has not elapsed to permit the recovery of the ventricular muscle from its refractory state. The latter, except for its time relations, is approximately normal. An auricular extrasystole, therefore, causes a premature contraction of the whole heart. The premature auricular beat prevents the occurrence of the next normal auricular impulse and the pause of the auricle which follows the abnormal auricular contraction is usually precisely equal to a normal interval. This fact has been explained upon the assumption that, normally, impulse formation in the S-A node is the

¹ It should be noted that except in the case of interpolated beats there is not an extra or additional beat, as the term extrasystole seems to imply. The premature beat, in effect, displaces the normal beat.

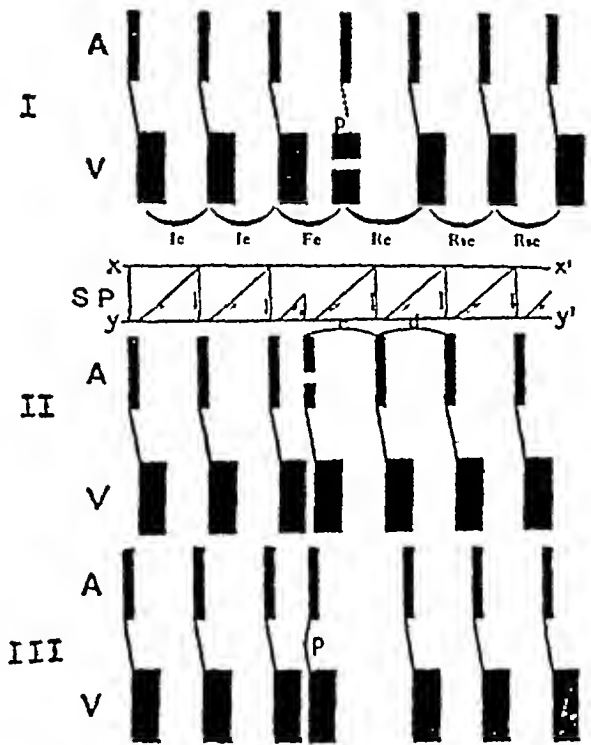


FIG. 24.4 (After Lewis) I A diagram illustrating disturbance of the heart's mechanism when a systole is caused by exciting the ventricle during diastole. Ic, initial cycle, Fe, forced or extrasystolic cycle, Rc, returning cycle, and Rsc, restored cycles. p is the premature or forced beat. Note that the auricular rhythm remains undisturbed. The forced and returning cycles are together equal in length to two initial cycles. II A diagram illustrating the events when a premature contraction is excited from the pacemaker. Stimulus production in the tissue which originates the heart rhythm is indicated by the line SP, the impulse is supposed to discharge when it reaches the line xx' and to fall at each contraction of the heart to the level yy'. c and d are equal in length. III A diagram illustrating a premature beat arising in the A-V node.

result of the liberation of energy which has been built up during the previous quiescent period. Upon the occurrence of the abnormal impulse this store of energy, accumulated for the normal impulse, is discharged and a definite time interval must elapse before it is again built up to the required level (see fig. 24.4, II). Sometimes, how-

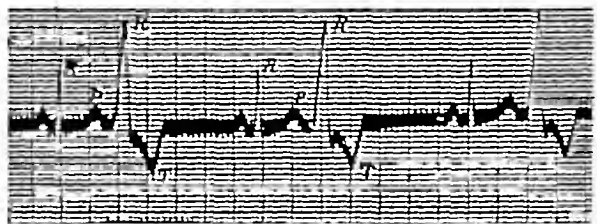


FIG. 24.5 Ventricular extrasystoles (bigeminal pulse, p. 233) arising in right ventricle—left axis deviation (p. 224) (Kindness of Dr. John Hepburn)

ever, the interval following the premature contraction is slightly lengthened, suggesting that the rate at which the S A node builds up its store of energy is lowered. In any event, there is rarely a long (compensatory) pause following the beat of the *ventricle*, the interval between the two normal beats, i.e., from the beat preceding to the one succeeding the premature contraction, being nearly always shorter than two normal cycles. In other words, the normal auricular impulse following the premature auricular contraction upon reaching the ventricle does not "miss fire", as in the case of ventricular extrasystole, but calls forth a response from the ventricle.

(3) Nodal extrasystoles

Extrasystoles occur sometimes as a result of impulse formation in the A-V node or supraventricular part (stem) of the bundle. On account of its central position between the two chambers, impulses arising in the A-V node pass upward and downward to cause simultaneous or nearly simultaneous responses from auricle and ventricle. Sometimes the ventricular contraction may actually occur first, in which case it is suggested that the impulse has arisen in the stem of the bundle and so has its course to the ventricular muscle considerably shortened. The extra cycle is usually, though not invariably, followed by a compensatory pause (fig 24.4, III).

In rare instances extrasystoles arise as a result of abnormal impulses initiated in the sino auricular node—*sinus extrasystoles*. Except for the interposition of the extra beat the rhythm of auricles and ventricles is but little disturbed. The interval following the extra beat is normal in length or slightly shortened.

The effects of extrasystoles upon the characters of the electrocardiogram and of the arterial pulse

THE ELECTROCARDIOGRAM In *ventricular extrasystoles* the electrocardiogram shows irregularity in the spacing of the ventricular complexes. The following characteristics are found:

(1) The intervals between the R wave caused by the premature beat and the corresponding waves of the normal beats preceding and following it, respectively, are altered in length. The interval between the last normal R wave and the premature R wave is short, while the interval following this to the next normal R wave is prolonged—compensatory pause. The time elaps-

ing between the two normal R waves is usually equal to the length of two normal cycles.

(2) The premature R wave is not preceded by a P wave. Since the premature ventricular contraction occurs unrelated to auricular systole it frequently happens that a normal contraction of the auricle occurs about the same time as the ventricular extrasystole. P and R waves then become fused. At other times the P wave follows closely upon the premature R wave.

(3) The P waves are equally spaced and some appear which are not succeeded by a ventricular complex (refractory period of the ventricular muscle). In the case of the *interpolated* type of extrasystole, however, each P wave is followed by an R wave, and no long pause is seen.

(4) Ventricular extrasystoles also show abnormalities of the QRS complex which distinguish them from premature contractions of auricular or nodal origin. An impulse arising in the heart below the point where the bundle forks will activate one ventricle slightly in advance of the other. It is to be expected then that the QRS deflections of the electrocardiogram will be a record of unbalanced electrical effects. This is actually the case (fig 24.5). In other words, if the extrasystole arises in the left ventricle its record will be deformed much in the same manner as that already described (p 229) for right bundle branch block (right axis deviation—main initial deflection downward in lead I and upward in lead III). If the premature beat arises in the right ventricle the electrocardiogram will show the features of a left branch defect (left axis deviation—main initial deflection upward in lead I and downward in lead III).

In *auricular extrasystoles* the electrocardiogram shows disturbances in the timing of both the auricular and the ventricular complexes, but, as already mentioned (p 231), a normal or only a slightly lengthened pause follows the extra beat. The P waves are unequally spaced but each is followed by a ventricular complex. The abnormal auricular wave may coincide with and be buried in the QRS complex of the preceding normal cycle.

The records of *extrasystoles arising in the I-V node* or supraventricular part of the bundle are variable according to the timing of the auricular and ventricular contractions. When the auricles and ventricles are excited simultaneously the P and R waves become fused. When the two chambers are not activated simultaneously, the P pre-

cedes the R wave by a short interval, or the order of the waves may be reversed (R-P interval)

THE ARTERIAL PULSE It has been demonstrated that several long-recognized irregularities of the pulse are the result of extrasystoles. For example, in the irregularity known as *intermittent pulse* there appear from time to time relatively long intervals during which no beat is felt in the radial (fig 246, a). The intervals are most pronounced when a premature contraction of the ventricle which is too weak to open the semilunar valves occurs. This most commonly happens when the heart muscle receives the abnormal impulse during the earlier part of its relative refractory phase. The premature beat may be detected by hearing a faint first heart sound which is not succeeded by a second sound (p 211). No pulse is produced in the radial at the time, nevertheless the extrasystole may be followed by a compensatory pause, and graphic records show as a rule that the gap in the radial tracing is just equal to two normal cycles. That is, a beat is dropped completely from the arterial record. The detection of an extra contraction of the ventricle, however, enables the irregularity to be distinguished from the missed beats of partial heart block (p 227) which may give an arterial tracing with similar characters.

If ventricular extrasystoles which fail to open the semilunar valves are repeated after each normal beat, the long intervals separating the latter will cause pronounced slowing of the pulse rate. The pauses between the arterial pulses are doubled in length and the pulse frequency as a consequence is reduced to half the normal. Bradycardia produced in this way and sometimes termed *false heart block* is distinguished from true heart block by a study of the venous pulse or the electrocardiogram which reveals the extra ventricular complexes, or the faint sounds of the extra contractions may be heard upon auscultation.

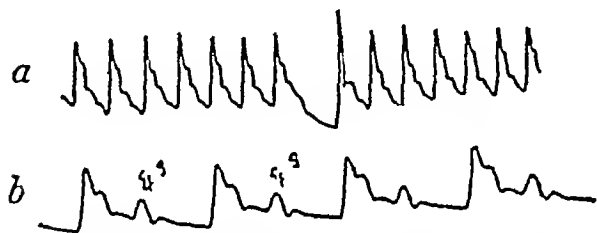


FIG 246 (After Price) a, intermission of the pulse, b, pulsus bigeminus, due to a single extrasystole with its succeeding compensatory pause occurring regularly after each normal beat. Ex S, extrasystole

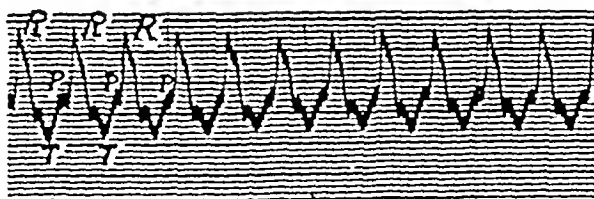
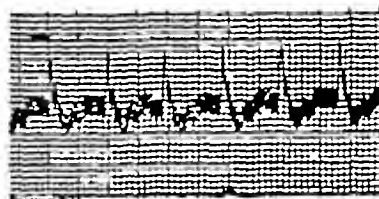
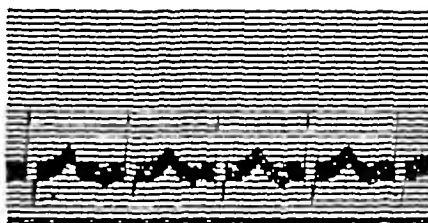


FIG 247 Top tracing, auricular tachycardia, middle tracing, nodal tachycardia, bottom tracing, ventricular tachycardia (Kindness of Dr John Hepburn)

When the extrasystoles are forceful enough to open the semilunar valves, and occur regularly one after each normal systole, paired pulse beats, each couple being followed by a long pause, are felt in the radial (fig 246, b). This type of pulse irregularity, which is sometimes seen following overdosage with digitalis, is called the *bigeminal pulse* (*pulsus bigeminus*).

II PAROXYSMAL TACHYCARDIA

This may be defined as a condition in which the rate of the heart is greatly accelerated for a longer or shorter period without obvious cause. The rate varies in different cases from 140 to 250 per minute. The onset of the paroxysm is sudden and the increased rate is maintained for a variable length of time with perfect regularity, successive cycles usually not varying in length by more than a hundredth of a second. The paroxysm lasts for only a few beats in some instances, in others it persists for a few minutes, hours or even days, though attacks of more than ten days' duration are very rare (Lewis). The attack ceases as abruptly as it commenced, the heart resuming its normal rate almost instantly.

The paroxysm, it is believed, consists of a series of rapidly recurring extrasystoles which completely submerge the physiological rhythm. The site of origin of the extrasystoles, as in the case of single premature beats, may be in the *auricle*, the

A-V node or stem of the bundle, or in the ventricle (fig 24.7) The auricular type is the most common, each auricular impulse spreads to the ventricle and causes a contraction whose features as indicated by the electrocardiogram are normal. The P wave is frequently inverted.

When the impulses arise in the A-V node or supra ventricular part of the bundle, the P-R intervals of the electrocardiogram are shortened. Inversion of the P waves is common. Or, the contractions of the two chambers may be simultaneous, the P waves being then buried in the ventricular (QRS) complexes. Again, the contraction of the ventricle may occur before that of the auricle, it then sometimes happens that a progressive lengthening of the intervals between the R and P waves (R-P interval) of the electrocardiogram is seen, ultimately an auricular beat is missed. This is termed *reversed heart block*. In other instances, as the R-P intervals reach a certain length, a contraction of the ventricle occurs prematurely, and is not followed by a contraction of the auricle. It is thought that the ventricular contraction is caused by the same impulse that caused the preceding auricular beat. That is, the impulse arising in the node first excites the auricle, then re-enters the junctional tissue, which has now recovered from its refractory phase, and passes downwards to the ventricle. This is spoken of as *reciprocal rhythm*.

When the impulses arise in the ventricle the QRS complexes have the characteristics of those caused by ventricular extrasystoles (p 230). The auricular rhythm is usually undisturbed but occasionally it is abnormal, for, when a series of rapidly recurring contractions arises in the ventricle the impulses may pass along the bundle in a retrograde fashion and activate the auricle to the exclusion of the normal impulse. In other words, the ventricle then sets the pace and the auricle follows. In such instances the P waves are inverted and succeed the QRS deflections, or are buried in the ventricular complexes as in the nodal type mentioned above.

III. AURICULAR FLUTTER

There are two forms of this disorder, *pure* and *impure flutter*. In both types the auricle beats at the phenomenally rapid rate of from 250 to 400



FIG 24.8 Pure flutter with a ventricular extrasystole (kindness of Dr John Hepburn)

beats per minute, but in pure flutter the rhythm is regular, in impure flutter it is irregular. Flutter differs from paroxysmal tachycardia in the following particulars:

(a) The auricular rate of beating is usually much greater than that seen in paroxysmal tachycardia.

(b) The disorder is of much longer duration, persisting unchanged for months or years, though it is sometimes transient.

(c) The ventricle fails as a rule to follow the rate of the auricle, a state of relative heart block becomes established as a result of the comparatively long refractory phase of the junctional tissue.

(d) It is produced by the passage of the impulse over one or more circular pathways—*circus movement* (p 235).

Auricular flutter may become converted to paroxysmal tachycardia.

In pure flutter the rhythm is remarkably regular (fig 24.8). The lengths of the auricular cycles vary no more than a few thousandths of a second over comparatively long periods. In impure flutter this constancy of cycle length is not seen and varying degrees of irregularity occur. In flutter the auricular walls do not completely relax. That is, though the proportion of active to inactive fibers varies during the rapid beating, at no one time are all the fibers in the relaxed state (p 206). In the electrocardiogram, therefore, the level of the tracing between the P waves lies for the most part above the isopotential line, only touching the latter for an instant. The level of the tracing is continually changing and the electrical changes of the auricle are carried over to alter the form of the ventricular complex. On the other hand, the fibers are never all contracted at the same time, so that auricular systole as well as diastole is incomplete.

The ventricle rarely keeps pace with the racing auricle, the refractory phase of the conducting tissue being longer as compared with that of the auricular muscle. As a consequence, a state of relative heart block develops and an auriculo-ventricular rhythm of 2 to 1, or less frequently, of 3 to 1 or 4 to 1, becomes established. Should the ventricle respond to each beat of the auricle, as occasionally happens, serious circulatory disturbances follow, ventricular diastole is so shortened that the ventricle receives a greatly reduced load of blood. The output of the heart may be so reduced as a result that loss of consciousness resulting in death may follow.

IV AURICULAR FIBRILLATION

This condition so far as fundamental causes are concerned may be looked upon as an advanced stage of flutter. The auricular muscle is the seat of incomplete contractions which recur at a frequency of from 400 to 600 per minute. So incomplete are the contractions and so rapidly are they produced that the individual beats are scarcely distinguishable from one another. The auricular cavity is never emptied of blood and its wall is a quivering sheet of muscle. Auricular fibrillation is the most common of all the serious cardiac irregularities, being associated, according to Lewis, in 60 to 70 per cent of all cases of cardiac failure in hospital practice. It is most frequently seen in mitral stenosis and in thyrotoxicosis (ch 58). It rarely occurs in the absence of myocardial disease.

Only a proportion of the auricular impulses pass through the A-V bundle and activate the ventricle. The relatively long refractory period of the conducting tissue shields the ventricle from the high rate of the auricular beating. The arterial pulse, nevertheless, is usually considerably faster than the normal (100 to 150) though it may be normal or even slowed. Those impulses which reach the ventricle do so in a somewhat haphazard manner, and indeed one of the most characteristic features of fibrillation of the auricles is absolute irregularity in the rate and force of the ventricular beats. These features are expressed in the terms "*delirium cordis*", "*complete irregularity of the pulse*" or "*perpetual arrhythmia*", which were applied to the condition before its true nature was recognized. A proportion of the heart beats are frequently so weak that they fail to cause a pulse in the radial. The apex beat is therefore much more rapid than the pulse. The former, for example, may be 150 and the latter only 60 or 70. The difference is called the *pulse deficit*. With treatment and improvement in the condition of the cardiac muscle the pulse rate therefore may increase.

The venous pulse in auricular fibrillation is of the ventricular form, *a* waves are absent, being represented by a series of rapid vibratory waves (*f* waves). Similarly in the electrocardiogram, small rapid undulations replace the *P* waves (fig 24 9).

THE UNDERLYING PROCESSES CONCERNED IN THE PRODUCTION OF FLUTTER AND FIBRILLATION OF THE



FIG 24 9 Auricular fibrillation, lead 2 (after Lewis)

AURICLES The *circus movement* theory versus the theory of numerous separate points of excitation. The observation that fibrillation could be induced in animals by electrical stimulation has led to a much clearer understanding of auricular fibrillation and flutter in man. Lewis, employing faradization as the method of inducing fibrillation, studied the subject intensively and, applying the results of the fundamental work of Mayer, of Mines and of Garrey conceived that the excitation wave followed a circular pathway through the auricular muscle—the so-called *circus movement*. That is to say, the wave starting at one place took a devious course through the cardiac musculature, returned to the point from which it started and re-entered the path which it had previously traversed. Mayer induced a circus movement in the umbrella of the jellyfish (*Medusa*) by creating a local block and applying a stimulus to one side of the blocked region (fig 24 10, I, A). The contraction wave which resulted was forced as a result of the block to take a unidirectional course, and after completing the circuit of the disc of tissue returned to the region of the block, which by this time had disappeared. If the tissue from which the wave had been initiated was again excitable, i.e., had passed from the refractory state, the wave circled the ring a second time, then a third time, and so on repeatedly. When, on the other hand, the disc was stimulated in the absence of a block, a contraction wave set out in both directions and the two waves meeting, after having completed half the circumference of the strip, were suppressed at B (fig 24 10, II). That is, further progress of the waves was arrested, for each came to a region of tissue which, being occupied by the other wave, was refractory.

Mines and Garrey linked up Mayer's observations upon *Medusa* with the fibrillation of cardiac muscle. Mines showed that a similar circular motion of the contraction wave could be induced in the cold-blooded heart. Muscular rings cut from the auricles of large rays were employed. Garrey thought that he had refuted

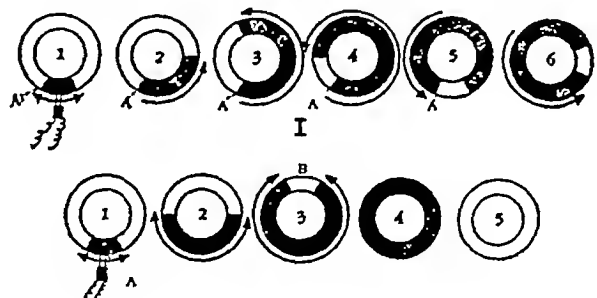


FIG 24 10 (Modified from Lewis) Black = contraction wave. Description in text.

the prevailing belief at that time, namely, that fibrillation of the auricle was due to the initiation of impulses from multiple foci throughout the auricular muscle—as proposed originally by McWilliam Garry found that if he cut a small piece from the fibrillating auricle, the severed tissue at once ceased to fibrillate, an event which he contended could not occur if numerous points of excitation existed, but such would naturally result if the impulses travelled over a circular pathway from which other parts of the auricle were excited. He induced a circulating contraction wave in large annular strips cut from turtles' ventricles.

Evidence was obtained by Lewis which seemed to show that the impulses in the fibrillating auricle travelled in the natural rings of muscle encircling the openings of the vena cavae.

The theory of circus movement has been very widely though not universally accepted. Bruns and Katz contended that if a circus movement is responsible for flutter and fibrillation more than one circulating wave must exist. They base this conclusion upon the observation that if after experimental flutter or fibrillation has been established, the auricles (or in the case of ventricular fibrillation, the ventricles) are separated from one another by crushing between them, the rapid beating persists with little change in each of the separated parts. Scherf, as a result of his failure to abolish auricular fibrillation by clamping or ligaturing the auricular muscle in the path of the supposedly circulating waves, was one of the first (1928) to throw doubt upon the existence of a circus movement. A little later Andrus and Carter concluded, from their experimental results in which fibrillation was induced by a single induction shock applied to the auricular appendix, that the waves did not originate and travel in a ring of muscle.

The most recent and convincing evidence against the circus movement theory has been secured by Prinzmetal and his associates. They induced auricular fibrillation by the application of aconitine to the auricles of dogs and observed the auricle by means of high speed cinematography. They saw no indication of a circus movement. Two types of activity were observed, (a) contractions and relaxations of microscopic or minute segments of muscle, which they term the "M" activity, and (b) stronger contractions and relaxations of macroscopic or larger muscle segments which appear upon the background of the innumerable "M" contractions. This type of movement they have designated the "L" activity. Both types of activity are heterorhythmic and utterly chaotic, and apparently arise from numerous discrete ectopic foci.

The actions of the cardiac glycosides, such as digitalis, and of quinidine upon fibrillation and flutter, and in heart failure

DIGITALIS belongs to a class of drugs known as the cardiac glycosides, which also includes strophanthin

ouabain (G-strophanthin from *S. gratus*), and squill. Digitalis in therapeutic dosage exerts a three fold effect upon the action of the heart. (a) It increases the force of the ventricular contraction both of the failing (hypodynamic) heart, and of the normal myocardium as shown for the isolated papillary muscle of the cat by Cattell and Gold. This is the drug's most valuable property. Ouabain and strophanthin have a similar action. (b) It depresses A-V conduction, and (c) it stimulates the vagus, a reflex effect through the carotid sinus and the vagus center (Heymans and associates). This action is seen only in the failing heart, not in rapid heart action in the absence of failure.

These last two effects cause slowing of the ventricle in fibrillation and flutter, but exert no significant effect upon the rapid auricular rate of impulse production. The depression of A-V conduction shields the ventricle from the fibrillating auricle, the weaker impulses fail to reach the ventricle, the pulse therefore becoming slower and more regular. In therapeutic dosage the drug has little or no effect upon the rate of the heart if the auricular rate is normal. In overdosage the drug may cause complete blockage of impulse conduction in the A-V bundle. It used to be thought that digitalis was of value only in fibrillation or flutter, but it has been shown conclusively within the last two decades that it strengthens the ventricular beat, and is of the utmost value in cardiac failure whether or not either of these arrhythmias are present. Indeed it is of little value in fibrillation or flutter unless myocardial failure exists as well.

Certain effects of the cardiac glycosides upon the metabolism of the myocardium have been demonstrated which help to explain their beneficial action upon the cardiac contraction. (a) They increase the oxygen consumption of the heart. (b) They increase the utilization of glucose and of lactic acid. (c) They cause a partial restoration of the adenosinetriphosphate and phosphocreatine contents of the failing myocardium in which these energy rich phosphate bond metabolites are depleted. In overdosage these drugs themselves cause depletion to the extent of 50 per cent or more of these materials.

The action of digitalis and other cardiac glycosides appears to be mainly in improving the utilization of chemical energy by the myocardium, rather than by increasing the total amount of energy liberated, though the latter probably also occurs.

The cardiac glycosides exert no beneficial effects upon myocardial insufficiency due to anoxia, in thyrotoxicosis, or in vitamin B₁ deficiency. The reason for this is unknown.

QUINIDINE (an isomer of quinine) also exerts a three-fold action in fibrillation. (1) It depresses or abolishes vagal tone and so lengthens the refractory period of the auricular muscle and decreases the transmission rate. This action upon the vagus is therefore opposite to that of digitalis. (2) It acts directly upon the auricular and

ventricular muscle, lengthening the refractory period (by from 50 to 100 per cent) and slowing the transmission rate. (3) It depresses conduction in the junctional tissues—a direct action. The abolition of the fibrillation of the auricle and the restoration of the normal rhythm was believed, on the basis of the circus movement theory, to be due to the lengthening of the refractory period of the auricular muscle, and as a consequence, to closure of the gap of excitable tissue between the crest and tail of the wave (see fig. 24-10). In other words “the head catches up to the tail” (Osler). But however fibrillation is produced, we know at any rate that quinidine restores the normal auricular rate and thus acts upon this arrhythmia in a manner different from that of digitalis. In the restoration of the normal auricular rhythm by quinidine fibrillation is frequently converted first to flutter.

The different effects of quinidine upon the heart interact in a complicated manner. For example, its action upon the ventricular rate will be the resultant of the following three effects:

(1) Rapid auricular beating tends automatically to depress conduction through the A-V connections. Therefore when the rate of the auricle is reduced by the drug, A-V conduction is increased.

(2) The reduction in vagal tone also increases A-V conduction.

(3) The direct effect of the drug upon A-V conduction, as stated above, is one of depression.

As a matter of fact (1) and (2) frequently overbalance the last effect (3) and some increase in the ventricular rate results.

Quinidine in overdosage produces severe toxic effects among which are auriculo-ventricular block, extra-systoles, paroxysmal tachycardia, and even death as a result of ventricular fibrillation. The return of the normal auricular contractions under quinidine treatment is sometimes, though rarely, followed by the dislocation of an intra-auricular thrombus and death from embolism. Complete standstill of the heart has also been reported as a result of the paralysis by the drug of the sino-auricular and auriculo-ventricular nodes and other tissues capable of impulse initiation.

V VENTRICULAR FIBRILLATION

The ventricular muscle may pass into a state of rapid, tremulous and ineffectual contractions closely similar in nature to the condition just described as occurring in the auricle. In animals, ventricular fibrillation may be initiated by direct electrical stimulation of the ventricular muscle, as was first shown by Ludwig in 1850. Mechanical stimulation of the ventricle, especially by pricking the tissue in the A-V groove, ligation of a coronary artery (Porter) or certain chemicals and drugs in excess, e.g., digitalis or calcium chloride, may

induce fibrillation. Levy found that chloroform anesthesia renders the hearts of experimental animals (cats) highly susceptible to fibrillation. A mere touch of a finger or instrument, the stimulation of a sensory nerve, section of the vagi or their paralysis by atropine, may set the ventricle fibrillating. The heart behaves as though sensitized by the anesthetic and ready at the least provocation to fibrillate. Adrenaline was found to greatly enhance the effect of chloroform—a fact which indicates the danger of adrenaline administration while a subject is under the effect of this anesthetic. Fibrillation may also ensue spontaneously under chloroform and Levy found that the condition is more likely to supervene when the animal is passing from deep to light anesthesia. Cyclopropane, like chloroform, has the effect of sensitizing the heart to the action of adrenaline in inducing ventricular tachycardia and fibrillation, but these cardiac irregularities are rarely seen under ether anesthesia. Quinidine reduces the susceptibility to fibrillation during cyclopropane anesthesia.

The effects of ventricular fibrillation upon the circulation are incomparably more serious than those of the corresponding auricular condition. This is evident when the importance of the two musculatures in the dynamics of the circulation are compared (p. 206). In fibrillation of the lower chamber the propulsive force of its contraction is practically abolished and the circulation comes to an end, death follows within a few minutes.

From experiments upon animals it is believed that many instances of cardiac failure in patients under chloroform are due to fibrillation of the ventricles. This is generally considered to be of sudden onset, but it has been shown that when the condition is induced in animals by chloroform it is frequently ushered in gradually. A solitary extrasystole first occurs, which is followed after a time by coupling, tripling, and later by short runs of extrasystoles. Longer paroxysms of rapid beating follow. Finally as the tachycardia becomes more rapid it merges into fibrillation.

When fibrillation is induced suddenly as by electric shock its development may be somewhat different. According to Wiggers and his associates, only the first contraction is a true premature beat, those which follow are caused by re-entry of the excitation wave. These investigators, who induced ventricular fibrillation in dogs by the application of a single strong induction shock to the ventricle late in systole, recognize four stages in the development of fibrillation. In the

first or undulatory stage, which lasts for only a second or two, the contractions are rapidly repeated but do not follow the same course over the surface of the ventricle, the electrocardiographic deflections show considerable variability in form. In the *second stage of convulsive incoordination*, which lasts for from 15 to 40 seconds, the contractions are more frequent and involve smaller areas of the ventricular muscle. The contractions of different areas of the muscle are out of phase so that the ventricle appears to be pulled about convulsively. The *third stage of tremulous incoordination* lasts for 2 or 3 minutes, the surface of the muscle is broken up into independently contracting areas of ever-decreasing size which are out of phase with one another. Thus, a tremulous appearance is given to the ventricles. The *final stage of atonic fibrillation* develops when the developing anoxia of the cardiac muscle causes weakening of its contractile force. This stage appears usually within from 2 to 5 minutes following the first stage and is marked by weak contractions or wavelets which travel only a short distance over the ventricular surface. It ends in complete cessation of all activity.

Ventricular fibrillation in man may result from —

- a) Electric shocks—electrocution, lightning stroke
- b) Chloroform or cyclopropane anesthesia
- c) Coronary occlusion and other causes of severe anoxia.
- d) Trauma to heart or chest wall
- e) Ventricular paroxysmal tachycardia, in which fibrillation may be a terminal event.
- f) Toxic doses of digitalis or quinidine
- g) Various diseases during the death agony

The fundamental factor or factors leading to fibrillation of the ventricles have been the subject of research by several investigators. Kebar and Hooker have found microscopical tissue changes in the dog's heart, in which fibrillation was induced by electric shock. There also occurred an increase in potassium in the outflowing fluid perfused through the heart. They attribute the fibrillation to a leakage of potassium from the cells and an unbalance of the potassium ion. The addition of potassium to the perfusion fluid brings the fibrillation immediately to an end and restores the normal beat. According to Nahum and Hoff, the essential condition for the onset of fibrillation is the

establishment of a block or blocks of the conducting system in a heart whose automaticity is simultaneously stimulated, either factor alone is ineffective. They found that the rapid injection of a concentrated solution of KCl throws the ventricles immediately into fibrillation. This does not occur if the injection is made slowly, for then automaticity instead of being stimulated is depressed, block of the conducting system occurs but fibrillation does not ensue. Thus, they believe that the two essential elements in the production of ventricular fibrillation are intraventricular block and automaticity of the ventricles. Ventricular fibrillation, though usually fatal is not invariably so, for rare instances have been reported in which the ventricles after fibrillating for a brief period resumed their normal rate and recovery occurred. Two methods, chemical and electrical, have been used in attempts to restore the normal rhythm to the fibrillating ventricles. Hooker has shown the efficacy of an excess of potassium in stopping fibrillation and of calcium in restoring the normal beat in the hearts of dogs subjected to electric shock. A 0.5 per cent solution of KCl is injected under pressure into the carotid toward the heart, so that it reaches the coronary system. This stops the heart. When a 0.023 solution of CaCl₂ is then introduced by the same route, the normal cardiac rhythm, in a successful experiment, is restored. Hooker and his associates showed that defibrillation of the dog's heart can be accomplished and the normal beat restored if a countershock, consisting of an alternating current of about one ampere, is passed through the heart. In fibrillation due to anoxia (e.g. caused by coronary occlusion) the heart may be defibrillated by either of these two means but the heart muscle is usually unable to develop a forceful contraction owing to the oxygen lack. Wiggers has modified Hooker's procedure by sending into the heart a series of shocks (3-7) of about one second duration and one or two seconds apart. He recommends that when fibrillation follows coronary occlusion cardiac massage should be practiced while the countershocks are being given, in order to increase the blood flow to the ventricular muscle. In fibrillation due to electrocution or other cause, the countershocks may be applied through the chest wall although much stronger currents (20-30 amperes) will be required. It is obvious, however, that any method devised for the resuscitation of the fibrillating human heart has little practical application on account of the short time which the heart survives, even though the necessary equipment were on the spot.

D ALTERNATION OF THE HEART PULSUS ALTERNANS

This is a condition in which every second wave in a pulse tracing is of relatively small amplitude. This peculiarity of the arterial pulse is due to alternate variations in the strength of the ven-



FIG 24 11 Radial tracing showing pulsus alternans (after Mackenzie)

tricular systoles, and to a smaller quantity of blood being ejected into the aorta during the weaker beat Figure 24 11 shows a typical sphygmogram of this condition There is as a rule little or no difference in the lengths of the intervals between pulse beats When a slight difference does exist, the interval succeeding the strong beat is then longer than that following the weak beat It will be remembered that in the bigeminal pulse, which might in some instances be confused with alternation, there is inequality in cycle lengths (p 233), but the longer interval follows the *weak* (premature) beat Furthermore, in alternation the ventricular rhythm does not share, or does so very rarely, in any irregularity of the pulse intervals which may occur in the arterial tracing The electrocardiogram, for instance, shows no discrepancies in the length of the intervals between the R waves The slight variations in the pulse intervals are attributed by Lewis to a slower rate of transmission of the weaker pulses to the periphery When records are taken simultaneously of the apex and the arterial pulse, it is sometimes found that the weak beats of the former coincide with the strong beats of the latter This discordance between apex and arterial beats is explained upon the assumption that those muscle fibers which contract during the weak apical impulses

though less numerous are actually more effective in ejecting the blood than those causing the stronger thrusts at the apex (see, Theory of Alternation, below) The appearance of alternation in the electrocardiogram (e g, alternate variations in the height of the R wave) is rare When this so-called *electrical alternation* does occur the larger deflections sometimes correspond to the weaker pulse beats But, it will be recalled in this connection that the deflections of the electrocardiogram are determined by the balance of the electrical forces developed during the cardiac contraction rather than upon their total value Electrical alternation may occur without alternation of the pulse

It is not possible to detect pulsus alternans by palpation of the pulse, the variations in strength of the pulse beat being too slight to be perceptible, but it is clearly revealed in the sphygmogram (fig 24 11) It may also be detected by means of a blood pressure armlet The pressure in the armlet is raised gradually, when it is found that at a certain level the weaker beats are suppressed, but the stronger beats get through The pulse at the wrist is then precisely half the ordinary rate The pressure during the weaker beats may be as much as 25 mm Hg below that during the stronger, but usually the pressure difference is not more than

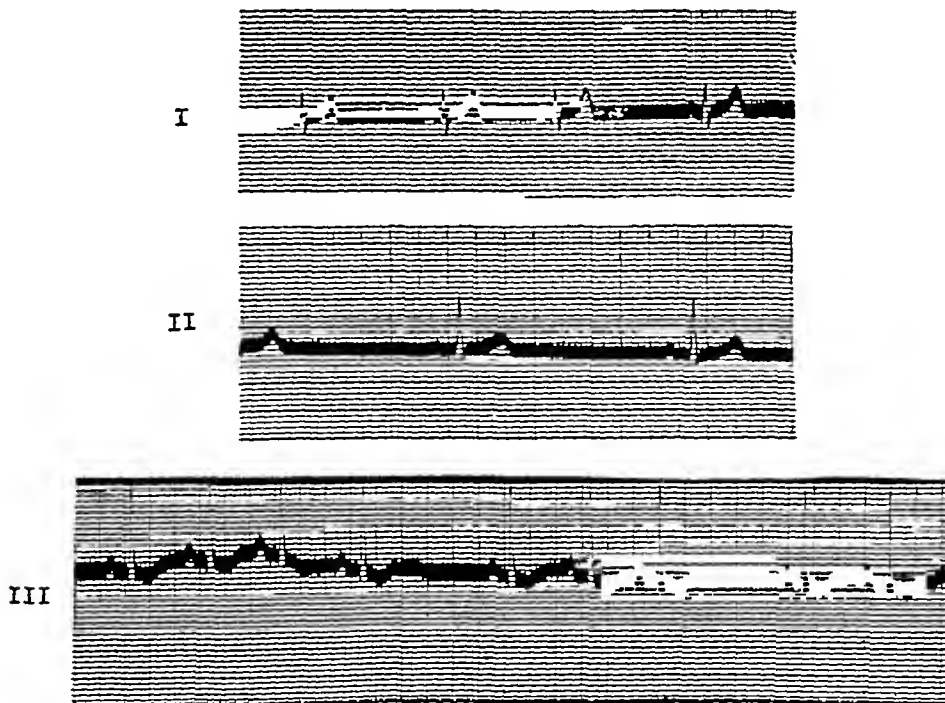


FIG 24 12 I, sinus arrhythmia, II, sinus bradycardia, III, sino-auricular block (Kindness of Dr John Hepburn)

5 or 10 mm Persistent alternation of the heart when the pulse is slow or of normal frequency is usually indicative of grave disease of the myocardium. Alternation sometimes also occurs at rapid rates of beating, e g, auricular fibrillation, paroxysmal tachycardia etc, but it is then of less serious significance.

THEORY OF ALTERNATION The "all or none" law states that the cardiac muscle responds maximally to any stimulus that is capable of evoking a response. It is apparently difficult to reconcile this principle with the varying force of the ventricular contractions in alternation, unless it be assumed that during the weak beat a smaller proportion of the heart fibers respond than during the strong beat. This is the prevailing view.

The heart, it is thought, must be in what has been called a *hypodynamic* state in order for alternation to occur. That is, the heart muscle is so depressed that only half of its fibers have recovered from the previous contraction before the impulse arrives. These fibers alone contract, and when the next impulse arrives they are still refractory, but those which had not previously contracted have recovered their irritability and now respond. In this condition all beats are weak but equal. Should a ventricular extrasystole occur in such a heart, the long pause which follows shifts the balance between the refractory and the non refractory fibers and precipitates the alternating rhythm. After the long pause of the premature beat a larger proportion of fibers have had time to recover and are therefore able to respond to the impulse. The next impulse, however, finds the muscle in a partial refractory state, i e., only a small proportion of the fibers have by this time recovered. They only can respond—a weak beat results. This small proportion of fibers when the next impulse arrives will in turn be refractory, but the larger proportion which had not previously contracted will now respond—a strong beat results. So the alternating rhythm is perpetuated, the fibers which responded during one contraction fail to contract at the next beat and conversely those which had failed to respond to one impulse respond to the next.

E IRREGULARITIES DUE TO VARIATIONS IN VAGAL TONE

I SINUS ARRHYTHMIA (fig 24 12, I)

This is a condition in which rhythmical variations in the rate of the whole heart occur synchronously

with respiration. It is due to alterations in the strength of the vagal influence upon the pacemaker (S-A node) as a result of the respiratory excursions, the heart rate increasing toward the end of inspiration and slowing toward the end of expiration. It is a youthful irregularity, being very common in children, and may be considered a physiological phenomenon. That it is entirely of vagal origin is shown by the fact that it is abolished by atropine. It also disappears when the heart rate increases as a result of exercise, fever, etc, but is enhanced by deep breathing.

II PHASIC IRREGULARITY

In this disorder periodic slowing of the heart occurs for a few seconds quite independently of the respirations. It also is a vagal effect since it is abolished by atropine. The manner of its production is unknown. It occurs in convalescence from acute fevers, and sometimes during the administration of digitalis.

III SINUS BRADYCARDIA (fig 24 12, II)

This is a persistent slowing of the whole heart due to increased vagal tone influencing the sino auricular node. The rate may be as slow as 40 per minute. Bradycardia of this nature occurs in apparently healthy persons, many of whom are athletes.

IV SINO AURICULAR BLOCK (fig 24 12, III)

The entire heart (auricles and ventricles) misses a beat at regular or irregular intervals. The condition thus differs from A V block in which only the ventricle misses (p 227). A complete set of waves is therefore dropped from the venous or electrocardiographic tracing, and the arterial pulse intermits. Since the condition is temporarily abolished by atropine and may be induced by stimulation of the vagus, it is probable that the missed beats are due to the action of the nerve upon the S-A node. Sino-auricular block sometimes results from digitalis administration.

V AURICULOVENTRICULAR BLOCK

Defective conduction between auricle and ventricle due to heightened vagal tone is occasionally seen, it is abolished by atropine.

CHAPTER 25

THE REGULATION OF THE HEART'S ACTION, HEART FAILURE

THE HEART RATE

In general, the rate of the heart bears an inverse relationship to the size of the animal, and a direct relationship to the metabolic rate. The heart rate in the canary, for example, is in the neighborhood of 1000 beats per minute, whereas that of the elephant is about 25. The average rate in adult man is around 70 per minute, but there is a rather wide variation between individuals, a rate considerably below or above this average being not uncommon. Muscular training tends to reduce the cardiac rate, athletes not infrequently having a pulse rate between 50 and 60. On the other hand, a rate between 80 and 90 is sometimes seen in other healthy persons. The rate diminishes progressively from birth, when it is around 130 per minute, to adolescence, but increases slightly again in old age. Among physiological conditions which temporarily increase the heart rate are *muscular exercise* (p. 246), *emotional excitement* and *high environmental temperature*. It also increases somewhat during *digestion*. The rate is lowered during *sleep* (55 to 60). Among pathological conditions which cause an increase in cardiac rate are *hemorrhage*, *surgical shock*, *hyperthyroidism*, *fever* (an increase of 10 beats per minute, per 1° Fahrenheit rise in temperature) and certain *cardiac arrhythmias*, e.g., *paroxysmal tachycardia* (p. 233), *auricular fibrillation*, etc.

Tachycardia and *bradycardia* are general terms used to denote, respectively, any considerable increase in heart rate above, or reduction below, the normal average.

Cardiac behavior is influenced by three agencies: I, *nerves*, II, *chemical materials*,—hormones, metabolites and inorganic salts in the blood, III, *mechanical effects* exerted upon the muscle fiber itself by the blood within the heart chambers. Each of these will be considered separately.

I NERVOUS CONTROL OF THE HEART

The heart, as we know, beats rhythmically after its complete separation from the central nervous system (p. 191) but in the intact animal this automatic action is under the continuous influence of nervous impulses. The nervous mech-

anism comprises groups of nerve cells in the medulla—the *cardiac centers*, various *afferent pathways* along which impulses are conveyed to these centers from numerous regions of the body, and the *vagus* and *accelerator* or *augmentor* nerves which transmit impulses from the centers to the heart.

THE VAGUS NERVES

The vagus nerves are cardio-inhibitory. This action was discovered by the Weber brothers in 1845. They convey fibers, belonging to the parasympathetic division of the involuntary nervous system, from a center in the medulla (*cardio-inhibitor center*) to the special tissues of the heart. The medullary center was located by Miller and Bowman in the dorsal nucleus of the vagus situated in the floor of the 4th ventricle. Weak electrical stimulation of this area produced slowing of the beat, and stronger currents, complete arrest of the heart. The cardiac fibers of the vagus separate from the trunk of the nerve in the neck between the origins of its superior and inferior laryngeal branches. Intermingling with fibers of the accelerator nerves they enter into the formation of the deep and superficial cardiac plexuses whence they are continued to the auricular muscle. Here they make connection with ganglion cells. Postganglionic fibers pass to the specialized tissue of the sino-auricular (sino-atrial) and auriculo-ventricular (atrio-ventricular) nodes where they are prolonged between the muscle fibers. They form a rich plexus and are seen to end as ring-shaped or club shaped structures ("boutons") upon the fibers of the specialized tissue. Many postganglionic fibers do not enter the nodal tissues but terminate in the auricular myocardium proper. Those which enter the auriculoventricular node do not extend beyond the upper part of the bundle. None has been found to terminate in the ventricular myocardium (Nomidez), or in the lower parts of the specialized conducting system.

It is probable that there are two types of vagal fibers, one type producing a change in cardiac rate (chronotropic effect), the other depressing

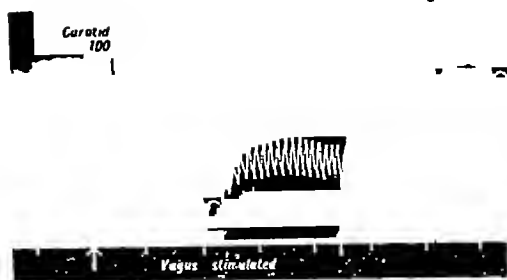


FIG 251 Showing the effect of vagal stimulation upon the arterial blood pressure. Note that although stimulation was continued, escape occurred, which in this instance was confined to the ventricles (McDowall)

conduction (dromotropic effect)¹ The S-A node is innervated mainly through the right, whereas the A-V node receives most of its parasympathetic fibers through the left vagus nerve

Stimulation of the cardiac vagus causes pronounced slowing or complete stoppage of the heart and, in consequence, a fall in blood pressure (fig 251). The slowing is brought about mainly through the lengthening of diastole, the duration of systole being relatively little affected. When the heart ceases to beat it does so in the diastolic phase of the cycle, appearing fully relaxed and engorged with blood. For these reasons the vagi are sometimes referred to as the diastolic nerves. The vagal effect upon the heart thus resembles the action of an excess of potassium (potassium inhibition, p 193).

When complete stoppage of the heart is caused by vagal excitation the ventricles, but usually not the auricles, commence after a time to beat again, though the stimulation of the nerve is continued. This is spoken of as the "*escape of the heart*" from the inhibitory influence. The amplitude of the returning contractions is frequently greater than usual. The escape phenomenon cannot be explained with entire satisfaction, it is not due simply to fatigue of the nerve and its failure to transmit impulses. The distention of the cardiac chambers (lengthening of muscle fibers, p 254) by the blood accumulated during the period of cardiac quiescence is probably a factor. There is some evidence that acceleration of acetylcholine destruction is concerned in the escape mechanism. Nahum and Hoff found that the escape phenom-

enon is greatly diminished after adrenalectomy and excision of the stellate ganglia, and conclude that adrenaline and the sympathetic nerves play an essential rôle in its production. The impulses which re-start the heart apparently arise from some level in the special conducting tissue below the level of the S-A node.

In mammals the vagus nerves exert their effects upon the heart through their action upon the auricular muscle and the junctional tissues. They exert no direct action upon the ventricular muscle. The effects upon the auricle are *slowing* of the beat (chronotropic effect) and *weakening* of its contraction (inotropic effect), *shortening of the refractory period* and a *rise in the threshold* (reobase) of excitation, i.e., in order to evoke a response from the auricle a stronger stimulus must be applied to the muscle during vagal stimulation than at other times. The chronaxie of the auricular muscle is moderately shortened. The slowing, or stoppage of the ventricular beat is an indirect effect due to auricular slowing, or arrest of the auricular contractions, or to depression of conduction in the auriculo-ventricular connections. That is, in the latter instance, partial or complete heart block is temporarily induced. If the A-V bundle is severed, vagal stimulation is then without effect upon the ventricular rate.

The slowing of the heart caused by vagal stimulation is often accompanied by an increase in amplitude of the beat as a result of the greater load of blood, and the consequent lengthening of the cardiac muscle fibers.

The effect of vagal stimulation upon the auricular rate is ascribed by Eyster and Meek to a shift of the pacemaker to a less rhythmical region, that is, to the production of a block in the upper part of the S-A node, or, with stronger stimulation, to complete suppression of impulse formation in this region. The duty of initiating the impulses devolves then upon regions, e.g., the lower part of the S-A node, the A-V node, or other parts of the special tissues possessing lower rhythmical powers. Evidence for this view is afforded by the observation that shortening of A-V intervals or the development of nodal rhythm (see p 232) may result from vagal stimulation (p 193). It has also been shown that primary negativity can no longer be located in the upper part of the sinoauricular node when the vagus (right) is strongly stimulated but is found to have descended to the lower part of the S-A node or to

¹ To the other effects of the vagus on the heart namely, on excitability and contractility, Engelmann, applied the respective terms, *batmotropic* and *inotropic*.

the A-V node Potassium salts also cause a shift of the region of primary negativity from the upper to the lower part of the S-A node—another example of the similarity between vagal and potassium inhibition Potassium also depresses auriculo-ventricular conduction

The right and left vagus nerves differ in their effects Stimulation of the right nerve from which, as already mentioned, the S-A node receives its main parasympathetic innervation, results chiefly in slowing and weakening of the auricular beat, and as a secondary result of this, in reduction of the ventricular rate Excitation of the left nerve, which ends chiefly in the A-V node, causes ventricular slowing or stoppage, by depressing A-V conduction and blocking the auricular impulses, little effect upon the auricular rate results An electrocardiogram, for example, taken during stimulation of the left nerve shows a greater number of P (auricular) than of R (ventricular) deflections

The difference between the effects of the two nerves varies considerably in different species For example, the left nerve in the tortoise depresses conduction from auricle to ventricle but exerts no effect upon the auricular rate, while slowing of the auricular rate is the sole result following stimulation of the nerve of the right side In the dog similar differences have been demonstrated Nevertheless, in the latter animal and in other mammals the right nerve depresses conduction to a greater degree than has been supposed, its effect being merely masked by the greater effect which this nerve, as compared with that of the left side, exerts upon the auricular rate Increase in heart rate automatically depresses A-V conduction, consequently when the right nerve is stimulated the slower rate which results induces an improvement in conduction which offsets any direct depressing effect of this nerve upon the A-V tissues Lewis and Cohn found that when arrangements were made to maintain the auricular rate constant during stimulation of the right vagus, the effect upon conduction was only slightly less than that shown by the left nerve

When the auricular impulse production is extremely rapid, as in auricular fibrillation (p 235), vagal stimulation increases the rate at which the impulse is transmitted through the auricle This effect, however, is not the result of any favorable influence which the nerve exerts upon the fundamental property of muscle-fiber conduction, but is due rather to the shortening of auricular systole and, in consequence, of the refractory period of the muscle The state of partial refractoriness which the rapid beating had induced is lessened This enables the impulse to take a shorter course which in turn causes a still more rapid fibrillating rate (up to

3000 per minute) This effect of stimulation of the vagus in the conditions of flutter and fibrillation has been termed *rapid re-excitation* The phenomenon is demonstrable in experimental flutter either by direct stimulation of the nerve or by the administration of a vagotropic drug such as pilocarpine It is possible that it occurs occasionally as a temporary manifestation in clinical flutter or fibrillation The effect of vagal stimulation upon the A-V connections during flutter or fibrillation is, as in the normal heart, to depress conduction and slow the ventricular rate

The tone of the vagus

During the normal life of the animal the vagus nerves exert a continuous restraint upon the action of the heart In other words the vagus, or rather the cardio-inhibitory center, possesses tone, impulses passing from it in a continuous stream to the heart This effect, which may be compared to the action of a dragging brake, can be readily demonstrated in animals by cutting or freezing the nerves The heart's action then immediately becomes greatly accelerated The increase in rate following the removal of the vagal influence also occurs though the stellate ganglia (ch 72) have been previously excised, the result therefore cannot be due to an increased action of the accelerator nerves The tonic action of the vagus nerves may be annulled by means of atropine, $\frac{1}{16}$ to $\frac{1}{8}$ grain being required in man to completely abolish their effects, the heart rate then increasing to 150 or 180 per minute The difference between this rate and the normal resting rate of 70 per minute, therefore, represents the vagal effect which is being constantly exerted under ordinary circumstances Various conditions, physiological and pathological, alter the tone of the vagus center The tone is naturally higher in some species, e g, the dog, which is capable of feats of endurance, than in others, e g, the domestic rabbit It also shows individual variations in man, athletes usually showing a higher tone than those who lead sedentary lives

Vagal tone is apparently reflex in nature and dependent upon afferent impulses flowing to the vagus center especially along the sinus and aortic nerves (p 246, 282) Section of these nerves causes an increase in heart rate and little further acceleration occurs as a rule when the vagi themselves are subsequently severed

THE ACCELERATOR OR AUGMENTOR NERVES

The accelerator fibers were described by Von Bezold in 1863 They belong to the thoracico-

lumbar division of the involuntary nervous system and arise from cells situated in the lateral horns of the upper thoracic segments of the spinal cord —I to V_T (see p 1094) These cells constitute a spinal cardio-accelerator center The preganglionic fibers enter the ganglionic cord of the sympathetic to connect with cells in the *inferior*, *middle* and *superior cervical ganglia* In many animals and also often in the human subject the inferior cervical and the first thoracic ganglia are fused into an irregularly shaped structure called the *stellate ganglion* from which accelerator fibers pass directly to the heart The heart also receives accelerator fibers *directly* from the sympathetic chain as far down as the 4th or 5th thoracic ganglion In order, therefore, to remove all accelerator influence from the heart it is necessary, as shown by Cannon, Lewis and Britton, and by others in man, to interrupt these connections as well as to remove the stellate ganglia The axons of the cells of the cervical ganglia (post-ganglionic fibers) form the *inferior*, *middle* and *superior cardiac nerves* (fig 25.2) These fibers, especially those forming the nerves of the right side, terminate in the sino-auricular node Those of the left side are distributed mainly to the A-V node and bundle According to Nonidez, the sympathetic efferent fibers which reach the heart are contained mainly in the middle cardiac nerve The superior cardiac nerve is distributed to the large arteries at the base of the heart while the

inferior cardiac nerve is mainly afferent. The spinal accelerator center is subordinate to higher centers The precise locations of the latter are not known, but the experiments of Beattie, Brow and Long indicate the presence of a center in the posterior hypothalamic region (ch 67), and Green and Hoff observed changes in heart rate, in blood pressure and in limb and kidney volumes upon stimulating the cerebral cortex (motor and premotor areas) in cats and monkeys A medullary center also probably exists

Stimulation of the cardiac accelerator nerves causes quickening of the rate of both auricles and ventricles and an increase in the force of the contractions We have seen that the vagus exerts its influence chiefly upon the auricle and junctional tissue, affecting the activity of the ventricle only indirectly The accelerator nerves, on the other hand, exert, as well, a direct action upon the ventricular muscle Stimulation of the accelerator nerves may so excite the ventricles as to induce fibrillation Fibrillation of the auricles, on the other hand, is not induced by this means We have also seen that adrenaline (which has an action upon the heart similar to that following stimulation of the accelerators) is particularly prone, during chloroform anesthesia, to induce fibrillation of the ventricles (p 237) The increase in the ventricular rate that occurs when these nerves are stimulated is accomplished usually through the shortening of both systolic and diastolic periods, but systole is curtailed to a greater degree than diastole, so that the ratio of the length of systole to the length of the entire cardiac cycle ($\frac{\text{systole}}{\text{cycle length}}$ ratio) is decreased

There is evidence that the effect of the accelerators upon A-V conduction is, as one would expect, opposite to that exerted by the vagus

Accelerator tone

The accelerators also exert a tonic action This has been demonstrated by excision of the stellate ganglia when the heart rate is reduced Gasser and Meek, for example, found that when the ganglia were removed but the vagi left intact, an immediate and marked fall in rate (about 40 per cent) occurred, further slowing occurred later which was attributed to a rise in vagal tone Slowing of the rate is produced, however, by excision of the ganglia, even though vagotomy has been performed previously, this fact of course precludes the possibility of the immediate reduc-

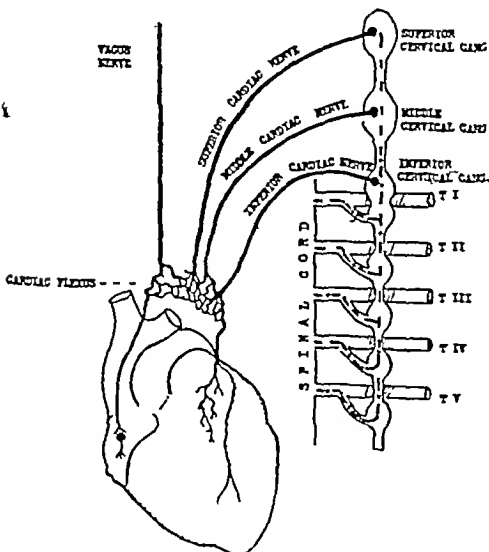


FIG 25.2 Diagram of the cardiac nerves Broken lines = preganglionic sympathetic fibers

tion in rate following excision of the stellate ganglia being due to increased vagal tone Chapman and associates observed slowing of the heart in man after bilateral excision of the sympathetic ganglia from the 2nd to the 5th thoracic ganglia inclusive In some instances a reduced cardiac response to exercise was observed after this operation

Bronk and his associates have demonstrated the existence of accelerator tone in the cat by recording the action potentials from nerves leaving the stellate ganglion A fairly continuous discharge of impulses at a rate of from 5 to 20 per second was observed Stimulation of the central end of an afferent nerve (e.g., sinus or aortic nerve) caused a discharge of impulses in the efferent fibers of the same frequency as those in the afferent impulses

CARDIAC REFLEXES

Under ordinary conditions, the activities of the cardio-inhibitory and cardio-accelerator centers which result in the continuous discharge of impulses along the corresponding cardiac nerves are in turn dependent to a very large extent, if not entirely, upon the reception of impulses by afferent paths In other words, the maintenance of the tone of the centers, and so of the normal resting rate of the heart, and the alterations in rate which occur under various physiological conditions are in large measure either reflex in nature or due to impulses received from cerebral centers The impulses which stream into the nervous centers arise in all parts of the body, the heart itself included By these influences the tone of either center may be exalted or depressed, and corresponding changes produced in the cardiac rate For example, Goltz showed many years ago that reflex inhibition of the frog's heart could be induced by repeated gentle taps upon the intestines Reflex slowing of the pulse can usually be demonstrated in the human subject by pressure upon the eyeball at the outer canthus (oculo-cardiac reflex), or by the stimulation of nasal branches of the fifth nerve Stimulation of afferent fibers in the respiratory passages² as by the inhalation of irritating vapors, e.g., anesthetics, is particularly likely to cause reflex inhibition of the heart Extrasystoles and bradycardia have been demonstrated electrocardiographically in man during abdominal operations, the irregularities being the consequence,

apparently, of visceral stimulation Excitation of the central end of various peripheral nerves, e.g., the sciatic, causes reflex changes in the pulse rate In these last instances acceleration is more readily obtained than inhibition The irradiation of impulses on to the cardiac centers from the cerebral centers, e.g., from the motor area at the commencement of muscular exercise (p 246) or from regions concerned with emotional manifestations, are held responsible for the changes in pulse rate which occur under these conditions

Afferent endings in the heart and aorta

Afferent fibers are contained in the cardiac vagus itself, the receptors of these lie within the heart tissues, and upon the aortic arch If, therefore, the cardiac vagus on one side be cut and its central portion (i.e., the end leading to the brain) stimulated, a reflex through the cardio-inhibitory center and the opposite vagus occurs which alters the cardiac rate The nature of the change in rate—whether acceleration or inhibition—which will result from stimulation of the central end of the vagus or of most other afferent nerves cannot always be foretold As a general rule, however, stimulation of afferent spinal nerves causes acceleration, and of afferent vagal fibers, inhibition The effect which is obtained depends also to a large extent upon the intensity of the stimulus, strong stimuli being more likely to increase the cardiac rate, weak ones to depress it It depends also upon the species, in some animals quickening is more readily obtained than slowing, in others the reverse is true For this reason it has been thought that the vagus and other nerves contained two types of afferent fiber, each type being responsive to different strengths of stimulation, and each being contained in variable proportions in different nerves as well as in the corresponding nerves of different species

In some animals the vagus gives off a branch composed entirely of afferent fibers This nerve terminates in the wall of the aortic arch and in the heart itself It is known as the *aortic* or *cardiac depressor nerve* (see p 282) When stimulated, this nerve induces cardiac slowing and vasodilatation, and, as a result, a fall in blood pressure

Marey's law

Marey's law states that the pulse rate is inversely related to the arterial blood pressure—a rise or a fall in pressure causing, respectively, a decrease or an increase in heart rate These ad-

² The rabbit's heart is slowed to between 50 and 60 beats per minute by the inhalation of ammonia By means of this response, it has been found possible to make slowing of the heart a conditioned reflex to the ringing of a bell (see Chapter 69)

justments are subserved by

(1) A reflex whose afferent limb is constituted of afferent vagal fibers (aortic nerve) ending in the aortic arch and heart.

(2) A reflex in which the *sinus reræ* forms the afferent limb. These mechanisms will be considered in chapter 27.

A fall in blood pressure causes an increase in the rate of the heart, however, even after it has been completely denervated, the acceleration is attributed to the liberation of adrenaline.

The Bainbridge reflex

The increase in heart rate induced by a rise in the pressure of blood entering the right auricle is known, after its discoverer, as the Bainbridge reflex. It is carried out presumably through the afferent vagal terminations beneath the endocardium and in the walls of the great veins near their entrance into the auricle. The nerve filaments are stimulated by the increased venous pressure which rises only after the cardiac chambers have been completely filled. Through this mechanism, the heart rate is adjusted automatically to the volume of blood poured into its chambers (the venous inflow).

THE RECIPROCAL ACTIONS OF THE CARDIAC CENTERS

Changes in heart rate are brought about not simply by an increase or a decrease in tone of one or other cardiac center but by reciprocal variations in the tone of both. An increase in heart rate, for example, is due to a decrease in tone of the cardio-inhibitory center accompanied by an increase in tone of the accelerator center. A decrease in heart rate is the result of increased tone of the cardio-inhibitory center together with reduced tone of the accelerator center. Such a balanced mechanism must make for more rapid, smoother and nicer adjustments of the cardiac rate than would be possible otherwise. Generally speaking, however, changes in tone of the cardio-inhibitory center appear to exert a somewhat greater influence than do alterations in accelerator tone. For example, the slowing of the heart which results from a rise in arterial pressure is much less pronounced if impulses from the cardio-inhibitory center have been prevented from reaching the heart by section of the vagi. After removal of the stellate ganglia, on the other hand, the cardiac response to a rise in blood pressure is reduced to a less extent. Also the increase in cardiac rate

which occurs at the commencement of muscular exercise is due to a reduction in vagal tone rather than to an increase in accelerator tone. The reciprocal relationship between the cardiac centers in the reflex acceleration of heart rate resulting from stimulation of the central end of the sciatic nerve has been demonstrated by Hooker and by Bainbridge. Though reduction in vagal tone was the more important factor in the accelerator response, increased tone of the accelerators was also shown. When, for example, the vagi were cut and the heart maintained at a constant rate by stimulation of the peripheral end of one of the cut nerves, or by the administration of pilocarpine, excitation of the central end of the sciatic nerve caused cardiac acceleration. This was attributed to an increase in accelerator tone, adrenaline liberation was not a factor since acceleration occurred after adrenalectomy. More recently, Moore and Cannon have found that after sympathectomy, vagal action is capable of slowing the heart from a rate of 150 beats per minute to 75, and the sympathetic mechanism, after vagotomy, can increase the rate from 125 to 225.

VOLUNTARY ACCELERATION OF THE HEART RATE. Several instances (Taylor and Cameron, and Ogden and Shock) have been reported of individuals possessing the power of voluntarily accelerating the heart rate. In one such case the effect was brought about apparently through the discharge of impulses along accelerator nerves since other sympathetic manifestations, e.g., vasoconstriction, glycosuria and dilatation of the pupils accompanied the increased pulse rate.

THE ACCELERATION OF THE HEART IN MUSCULAR EXERCISE AND IN EMOTIONAL STATES

There are two stages in the cardiac acceleration which accompanies muscular effort, namely, the increase in rate which occurs immediately upon the commencement of the work, and that which develops subsequently and more gradually. The acceleration of the pulse at the outset of the exercise occurs too promptly to be the result of an increased flow of blood from the contracting muscles to the heart (Bainbridge reflex). Miss Buchanan has shown, for example, that with even very light exercise, namely, clenching the fist, the diastolic period immediately succeeding the commencement of the movements is shortened. Thus, before the exercise the length of the cardiac cycle was 0.82 second (pulse rate 73). The first cycle after starting the exercise was 0.67 second.

Subsequently the cycles shortened progressively to reach a value of 0.56 second, this corresponds to a pulse rate of about 107 per minute. Moreover, it has been shown by other observers that the initial acceleration does not occur if the muscles are moved passively or tetanized by direct stimulation.

The last mentioned observation indicates that the effect is not the result of a reflex elicited through receptors situated in the muscles themselves. It is concluded that the initial acceleration results from impulses arising in the motor areas of the cortex and overflowing on to the cardio-inhibitory center and depressing its tone. On the other hand, *if the circulation to the arm or leg is arrested* while the limb muscles are contracting the pulse accelerates and, provided that the limb's circulation remains occluded, the increased rate persists after the exercise has ceased (Alam and Smirk.) The effect under such circumstances is evidently due to a reflex from the contracting muscles. Its persistence after cessation of exercise until the circulation is restored and its disappearance thereafter, and also its association with the pain of intermittent claudication (p. 300) strongly suggests that the reflex is initiated from afferent nerve terminals stimulated by metabolites which have accumulated in the contracting muscles. This cardiac reflex, though interesting as a physiological phenomenon, is probably of very minor importance in the increased pulse rate in muscular exercise.

That depression of vagal tone rather than heightened accelerator tone is the responsible factor in muscular exercise is evidenced by the following facts: (a) the shortness of the latent period which is a characteristic of vagal effects, (b) the abbreviation of the cycles through shortening of the diastolic period, which was noted above, is also a criterion of lowered vagal tone (p. 243). Bowen also found that during exercise in man the increase in rate was brought about entirely by shortening of diastole. Indeed, systole was actually lengthened at the commencement of the work. Gasser and Meek, experimenting with dogs, found that removal of the stellate ganglia had little or no effect upon the cardiac acceleration occurring early in muscular exercise, whereas this primary increase in rate did not occur after vagotomy. They found, however, that after the initial stages of the work, increased accelerator tone contributed as well to the increased frequency. Diastole and systole were both shortened though the vagal effect predominated. The important factor in the delayed effect is no doubt the rise in venous pressure and the elicitation of the Bainbridge reflex (p. 246). Gasser and Meek conclude from their experiments that the chief function of the accelerator nerves is to maintain a steady frequency of the resting pulse and that their influence does not alter to any marked degree

under ordinary circumstances the duty of varying the heart rate devolving chiefly upon the vagal centers. The accelerators provide a factor of safety and, under circumstances of exceptional stress, e.g., strenuous muscular exertion or emotional states, reinforce the vagal effect. The liberation of adrenaline is an additional factor.

It is probable that in such emotional states as fright, anger and mental excitement, the accelerator nerves play a more prominent part in increasing the heart rate than does a reduction in vagal tone. Bond found that sudden startling of an animal, in which adrenaline liberation had been prevented, caused a much greater increase in heart rate if the vagi had been sectioned and the accelerators left intact than in converse experiments (accelerators divided and vagi intact).

THE EFFECT OF DRUGS UPON THE CARDIAC RATE

The efferent vagus pathway, as already mentioned (p. 241), consists of two links—the preganglionic and postganglionic neurons. *Atropine* acts by antagonizing the action of the cholinester which, according to the humoral theory of the transmission of nervous effects, is liberated at the postganglionic terminals. Atropine therefore depresses or, in full doses, completely abolishes vagal action, the effect upon the cardiac rate being the same as if the vagus nerves were cut. *Muscarine*, an alkaloid present in poisonous mushrooms (*Amanita*) has a diametrically opposite effect. Its action upon the cardiac rate imitates that of vagal stimulation, namely, to cause slowing and, finally, complete stoppage of the heart in diastole. The effects of *pilocarpine*, *physostigmine* and the ester of choline, *acetylcholine* (see below) are similar to those of muscarine, causing cardiac slowing or standstill (see also ch. 72). Their actions are antagonized by atropine. Minute doses of acetylcholine have been shown by McDowall to have an opposite effect to that produced by larger doses, they stimulate the heart by the liberation within the heart muscle of an adrenaline-like principle. *Nicotine* first excites the vagus but later acts like atropine, causing vagal paralysis and increased cardiac action. Its site of action is upon the ganglion cell or at the synapse of the latter with the preganglionic fiber.

THE HUMORAL TRANSMISSION OF VAGAL AND ACCELERATOR EFFECTS

Some few years ago Loewi performed a series of experiments which showed decisively that cardiac inhibition resulting from vagal stimulation

is due to the liberation of a chemical substance possessing an action similar to that of *acetylcholine* (see ch 72). It was also shown that the augmentation and increase in heart rate which follow stimulation of the accelerators are due to the action of an *adrenaline like substance*. His results have been confirmed by others. The experiments were performed as follows: two frog hearts were perfused with Ringer's solution. One heart (donor heart) was inhibited by vagal stimulation. The perfusion fluid of this heart when perfused through the second heart (recipient heart) caused inhibition of it also. The effect upon the second heart could have been due only to the presence of a chemical material of some sort in the perfusion fluid derived from the first heart, it must have been produced by vagal stimulation. The

action of the perfusion fluid upon the recipient heart was annulled by atropine. The cardio-inhibitory action of choline and its ester *acetylcholine* is also prevented by atropine. The perfusion fluid from a heart during vagal stimulation has been shown, moreover, to cause contraction of the stomach wall of the frog (Brinkman and Van Dam) and of a loop of intestine, effects which are also produced by *acetylcholine*. The "vagus substance" (Vagusstoff as it was called by Loewi) is also, like *acetylcholine*, readily destroyed by alkali (or by the esterase mentioned below) but is stable in acid media.¹

The method used by Bain to demonstrate the humoral transmission of vagal effects is shown in figure 25 3.

It was later demonstrated by Loewi that the heart muscle contained an enzyme—*cholinesterase*—which rapidly hydrolyzed the ester after its liberation into the relatively inactive choline and acetic acid, and so limited the duration of its action. The action of this esterase is inhibited by physostigmine (eserine). The effect of the latter drug in inhibiting the heart depends solely upon this action. That is, it prevents the hydrolysis of the *vagus* substance which therefore continues to exert its effect.

The liberation of a substance possessing an *adrenaline like* action was demonstrated by perfusion experiments of a similar nature. Perfusion fluid collected from one heart during stimulation of the accelerator nerves when perfused through a second heart accelerated its rate.

As mentioned on p 247, minute doses of *acetylcholine* cause a sympathomimetic substance to be liberated in the heart muscle. This substance, though having many of the actions of *adrenaline*, dilation of pupil, inhibition of intestine and of the pregnant uterus (cat), and being antagonized by *ergotamine*, differs from *adrenaline* in other respects. The *adrenaline like* principle has been obtained from extracts of heart muscle which has not previously been treated with *acetylcholine* (v Euler).

¹ Howell was the first to suggest a humoral mechanism underlying the cardio-inhibitory effect. He reported an increase in the concentration of potassium in the fluid perfusing an isolated mammalian heart during vagal stimulation. The resemblance of the action of potassium upon the heart to that produced by stimulation of the *vagus* has been pointed out on page 242. Howell proposed the theory that the nerve impulses caused the breakdown of some organic compound with the liberation of an inorganic form of potassium.

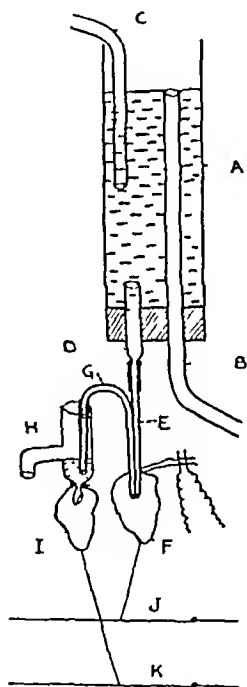


FIG. 25 3 The perfusion apparatus A is furnished with an overflow tube B, the height of which can be varied to allow alterations in the perfusion pressure. Fluid is delivered through the tube C. The fluid from A passes through the tube D to the inflow limb E of the double cannula which supplies the donor heart G. After irrigating the interior of the donor heart, the fluid passes by the outflow limb F to the glass cannulated tube H to which is attached the recipient heart I. This cannulated tube is provided with a lateral overflow, so that the pressure of the fluid supplied to the recipient heart remains constant. J and K are the levers to which the hearts F and I are respectively attached (Bain).

THE SENSORY NERVES OF THE HEART

The ordinary types of stimulus applied to the heart cause no painful or other sensation⁴ The exposed heart of a conscious human subject, for example, may be stimulated electrically or mechanically without any sensation of pain being experienced Pathological processes, especially coronary artery thrombosis and angina pectoris, may, on the other hand, give rise to acute and often agonizing pain (p 328) The results of experiments upon animals and of operations upon man, in which the cardiac sympathetic pathways have been interrupted or blocked by means of alcohol injections, indicate that the route for cardiac pain is solely through sympathetic afferents The pain fibers are contained mainly in the inferior cardiac nerves but some travel in the middle cardiac nerves The pain impulses then pass through the corresponding cervical ganglia, and the white ram and posterior roots of the upper four or five thoracic nerves (fig 25.2, p 244) Pain fibers also pass from the heart directly (i.e., not through the middle and inferior cardiac nerves) to the upper four or five thoracic ganglia Pain and stretch receptors have been described in the cardiac muscle

II CHEMICAL CONTROL OF THE HEART THE INFLUENCE EXERTED UPON THE HEART AND VASCULAR MUSCULATURE BY BLOOD-BORNE SUBSTANCES

The known materials in the circulating blood which influence the action of the heart are, (1) the *mineral constituents*—calcium, potassium and sodium, (2) *adrenaline*, (3) *acid metabolites*—carbon dioxide and lactic acid, and (4) *oxygen* The specific effects of the inorganic constituents upon the perfused heart have already been considered (p 192) Their proportions in the general blood stream do not vary under physiological conditions, and rarely in pathological states to a sufficient extent to affect materially cardiac behavior Adrenaline has a very definite action upon the heart and vessels which will be considered in Chapter 59

Carbon dioxide and lactic acid

Carbon dioxide exerts an effect upon the cardiac and vasomotor centers and directly upon the cardiovascular musculature Carbon dioxide

⁴ This fact was first shown by Harvey upon the exposed heart of the Viscount Montgomery (see p 597)

and lactic acid, formed during the activity of muscle and other tissues, dilate the peripheral vessels (p 291) The active tissues are thus supplied more generously with blood In muscular exercise the greater flow of blood through the muscles causes a greater mass movement of blood along the veins and a greater flow into the cardiac chambers The higher carbon dioxide tension in the venous blood coming to the heart from the skeletal muscles enhances the extensibility of the cardiac muscle fiber during diastole and, in consequence, exerts a favorable effect upon the filling of the heart The cardiac output is therefore increased (p 263)

The junctional tissues are particularly sensitive to high tensions of carbon dioxide, auriculo-ventricular conduction becoming markedly depressed when the carbon dioxide excess is such as to cause a fall in pH of the fluids bathing the cardiac muscle fibers At a pH of around 7.0 complete heart block occurs A less pronounced increase in CO₂ tension causes slowing of the heart, as a result of depression of the activity of the sino-auricular node as well as of increased tone of the cardio-inhibitory center The continued exposure of the heart to a high CO₂ tension causes weakening of the beat and the development of irregular rhythms

An abnormally low tension of carbon dioxide in the blood is sometimes referred to as *acapnia* (Greek *a*, privative + *kaphnos*, smoke) It occurs in conditions associated with increased pulmonary ventilation, the gas being "pumped" or "blown off" from the body in the expired air The carbon dioxide deficit produces effects which are the reverse of those caused by carbon dioxide excess Diastole is incomplete, the tone of the capillaries and small veins is increased (p 292), the venous pressure falls and the heart chambers are insufficiently supplied with blood The cardiac output, in consequence, is reduced Diastole becomes shorter and shorter with a consequent rise in cardiac rate, finally, as Henderson and his associates have shown in animals during prolonged artificial respiration with maximum ventilation, the heart enters into a condition of almost continuous contraction The blood pressure falls and death occurs from circulatory failure

Low carbon dioxide tension, if accompanied by a change in blood reaction toward the alkaline side, increases the rate of conduction over the auriculo-ventricular bundle The rate of impulse initiation

in the sino-auricular node is increased, the tone of the cardio-inhibitory center is lowered

Owing to the several factors involved, the effects of carbon dioxide deficit and carbon dioxide excess upon the action of the heart in the human subject offer many difficulties to the investigator. In forced breathing (which lowers the CO_2 tension of the blood) for example, the respiratory excursions, if violent, may interfere with the venous return and reduce the cardiac output. Vincent and Thompson observed a fall in blood pressure during forced breathing which they attributed chiefly to the mechanical effect of the respiratory movements upon the venous return, they assert that the lowered carbon dioxide tension played a minor rôle. If, on the other hand, the respirations are carried out at a different rate and depth, the venous return may be augmented and the cardiac output, in consequence, increased. The effect of forced breathing also varies with the subject and with the type of the breathing, thoracic breathing tending to increase the cardiac output, the abdominal type to reduce it. It is not surprising, therefore, that attempts by different workers to determine the effects of alterations in the carbon dioxide tension of the blood upon cardiac action have given conflicting results. Probably the most complete series of experiments relating to this question have been carried out by Grollman, who employed the acetylene method for the determination of the cardiac output (p. 269). He observed a greater output during forced breathing. He attributed the increase to the rise in metabolism incident to the extra work performed by the respiratory muscles, and to the mechanical effect of the respiratory excursions in augmenting the venous return. For, when the forced breathing was carried out with a *carbon dioxide-air mixture* which maintained the carbon dioxide percentage of the alveolar air around the normal value, an increase in the cardiac output occurred of the same magnitude as that occurring when the over-ventilation was carried out with ordinary air. Breathing a carbon dioxide rich mixture caused no change in cardiac output until the carbon dioxide in the inspired air reached about 6 per cent, then a marked increase in the pulmonary ventilation and a rise in the cardiac output occurred. The latter was attributed also simply to increased respiratory activity.

Grollman concludes that carbon dioxide lack must be of an extreme grade before it exerts an

effect upon the output of the heart, in ordinary experiments upon the human subject the loss of carbon dioxide falls short of the point where such an effect is produced. In disease, on the other hand, when a profound lowering of the carbon dioxide content of the blood occurs with a consequent change in the pH of the latter toward the alkaline side, deleterious effects upon the circulation and reduction in the cardiac output would be expected to occur as a result of incomplete diastolic filling. In order to prevent the "washing out" of carbon dioxide during artificial respiration Henderson advocates the addition of carbon dioxide (7 per cent) to the inspired air.

Oxygen

High tensions of oxygen tend to reduce slightly, and very low tensions to increase the output of the heart. In experiments with subjects who breathed mixtures of oxygen and nitrogen in varying proportions, Grollman found that the cardiac output did not alter until the oxygen content of the mixture had been reduced to 11.6 per cent. A moderate increase in the output then occurred. If the oxygen lack is profound or of long duration, the cardiac muscle suffers and the output of the heart then, of course, becomes reduced.

Reduction in the oxygen content of the arterial blood by 50 per cent or more causes a diminution in the amplitude of the T wave of the electrocardiogram. This wave may become flattened out or inverted in all three leads if the anoxemia persists. Changes resembling those caused by coronary infarction (ch. 28) may make their appearance. These electrocardiographic effects have been observed in disease and in airmen during ascents to over 5,000 feet.

Breathing a mixture with a very high percentage of oxygen slows the *heart rate*. In the early stages of oxygen lack the heart rate is increased. These effects are brought about presumably through the cardiac centers. In the later stages of anoxemia, heart block, extrasystoles and other cardiac irregularities develop and heart failure supervenes. The inability of the heart muscle to contract any considerable oxygen debt has been mentioned (p. 192).

Asphyxia

The combined effects of oxygen lack and carbon dioxide excess upon the heart rate are seen in asphyxia. At first, marked slowing occurs which

is largely a secondary effect of the rise in blood pressure (Marey's law, p 245) induced through the action of carbon dioxide excess and oxygen lack upon the vasomotor center (p 278). Later, as the strength of the cardiac muscle weakens, the blood pressure falls, the heart rate rises and finally cardiac irregularities appear. It has been shown by McDowall that mild asphyxia exerts a delayed effect upon the cardio-inhibitory center. If an animal is asphyxiated for a *short* period, cardiac slowing occurs after recovery. During the asphyxial period itself, slight cardiac acceleration may occur. The delayed effect is considered to be in the nature of an after discharge (ch 64) of the cardio-inhibitory center. From this observation it is suggested that the mild asphyxia which accompanies muscular exercise may be a factor in the development of the high vagal tone and slow heart rate found in athletes and others who are accustomed to perform heavy muscular work.

III THE MECHANICAL FACTORS CONCERNED IN THE REGULATION OF CARDIAC ACTION

The effects of mechanical or physical factors, e.g., venous inflow and arterial blood pressure, upon cardiac behavior have been studied in great detail by Starling and his associates. These classical experiments were carried out upon the heart-lung preparation of the dog. A description of the preparation follows.

The heart-lung preparation

This method enables the output of the left ventricle to be determined at the same time that any one of several factors, e.g., arterial resistance, venous inflow or temperature, is varied as desired, while the remaining factors are rigidly controlled. The following description is taken in the main from Knowlton and Starling's paper (see fig 254). The chest is opened and artificial respiration instituted. The arteries arising from the aortic arch are ligated, as well as the azygos vein. Cannulae are tied into the innominate artery (C.A.) and the superior vena cava (S.V.C.). The descending aorta is ligated. Blood which has been defibrinated or rendered incoagulable by the addition of heparin is placed in the reservoir F, this is connected by tubing of wide caliber with the cannula in the superior vena cava. The blood is warmed in its passage through a worm immersed in a water bath. The only channel open to the blood from the left heart is through the innominate cannula. The lesser circulation is undisturbed, the blood from the right heart is oxygenated by passing through the lungs to the left auricle. For the sake of simplicity the pulmonary

circulation is not shown in the figure. The innominate cannula connects by one limb of a T-tube with a mercurial manometer M^1 . The other limb of the T is connected by tubing with a second T-branch, v, the limbs of which are connected respectively with the resistance R, consisting of a rubber finger-stall enclosed in a sealed tube, T, and the inverted test-tube, B. The air in the latter, since it undergoes compression and decompression with the pulsatile variations in arterial pressure, acts like the air cushion in the stand-pipe of a water system, and serves in lieu of the elastic arterial wall (p 149). The resistance of the flow of blood through R can be raised to any desired height by means of the bulb S and the pressure bottle A. The pressure exerted upon the collapsible finger-stall is measured by the manometer M^2 . The blood, after passing through the peripheral resistance, collects in the tube N which is provided with a siphon tube Sy. This can be made of such dimensions that the blood is siphoned off when 10, 20 or 30 cc have collected. The change in pressure in the side tube leading to the tambour D, which occurs each time that the blood is siphoned off, causes a movement of the lever which is recorded upon a smoked surface. The output of the left ventricle over a given period is thus determined. The output of the left ventricle also gives the measure-

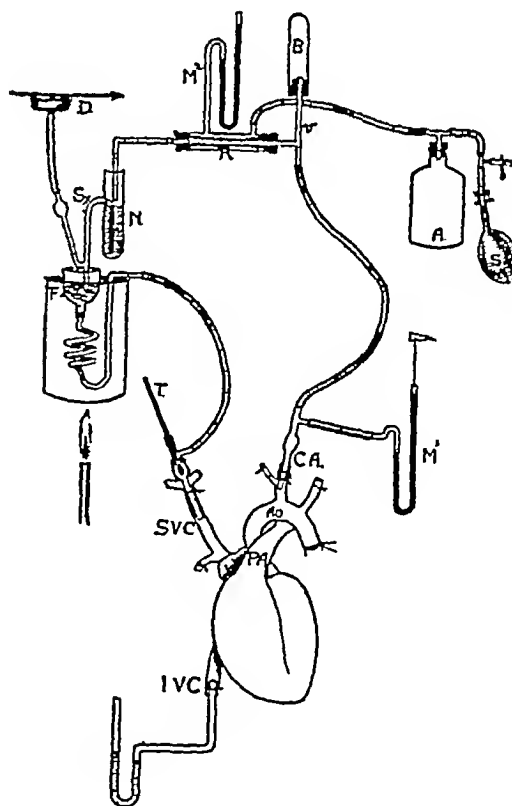


FIG 254 Arrangement of apparatus for heart-lung preparation of mammalian heart (Knowlton and Starling) Description in text

TABLE 19

Dog 5 15 kg Heart 56 gm. Response of heart to increased venous inflow (Patterson and Starling)

TEMPERATURE	ARTERIAL BLOOD PRESSURE	PULSE RATE PER MINUTE	PRESSURE INFERIOR VENA CAVA	OUTPUT PER MINUTE	OUTPUT PER BEAT	VENOUS SUPPLY
	mm Hg		mm H ₂ O	cc	cc	
34.5	100	126	20	560	4.2	Small
	110	126	60	1,100	8.7	+
	110	129	100	1,500	11.6	+
	96	132	150	1,720	13.0	+
	90	132	230	2,400	18.2	+
	90	135	250	3,000	22.3	Full

ment of that ejected by the right side of the heart, as well as of the venous inflow, since obviously the left ventricle can give out no more than it receives from the right, and must give out as much, so long as the circulation is maintained

Starling and his associates have shown that the dog's heart possesses a remarkable reserve power upon which it draws when confronted with a demand for greater work—increased venous inflow or raised arterial pressure. The heart was found to respond capably to an increase in the venous inflow of several hundred per cent, as the inflow increased a corresponding increase occurred in the output of the heart. The heart rate did not accelerate significantly, so the greater output was accomplished practically entirely by an increase in the quantity of blood ejected at each beat (see p 263). The results are shown in table 19. Figures for the venous inflow are not shown in the table, but the fact that the circulation was maintained shows that the heart was giving out as

TABLE 20

Dog 4.6 kg Heart 52 gm Venous supply maintained constant and arterial resistance varied (Patterson, Piper and Starling)

EXPERIMENT	BLOOD PRESSURE	VENOUS PRESSURE	OUTPUT	PULSE RATE
	mm Hg	mm H ₂ O	cc	
1	84	45	115	26.5
2	110	62	115	26.5
3	140	80	111	26.5
4	190	110	108	26.5
5	106	60	115	26.0
6	50	50	116	26.0
7	46	45	118	26.0

much blood as it took in—that is, the output was directly proportional to the diastolic filling. As the inflow increases, the venous pressure rises moderately and distends the heart cavities up to a certain optimal degree of dilatation. Beyond this point a falling off in output occurs. In other words, a point arrives when the heart is dilated beyond its physiological capacity, the muscle fiber is at a mechanical disadvantage and the heart fails to eject as much blood as it receives. The ventricle becomes distended with blood, and as a consequence, the venous pressure rises to an abnormal height. When, in the experiments of Patterson, Piper and Starling, the venous inflow was maintained at a constant rate and the arterial pressure raised by increasing the arterial resistance, there occurred a corresponding adaptation of the heart to the greater demands made upon it. The total output of the left ventricle remained practically unchanged though the arterial pressure was raised over 100 per cent (table 20).

It is quite evident from table 20 that the heart responded to a greater venous inflow by increasing its output per beat and not by increasing its rate. It is also true that, if the venous inflow is inadequate, i.e., the heart chambers are not completely filled at the end of diastole, or are just filled and no more, simply increasing the frequency of beat, as by raising the temperature, fails to increase the output. This is obvious, since if the beats have been occurring at the end of a period of rapid filling and no increase in venous inflow occurs as the rate increases, the acceleration in heart rate must be accompanied by a proportional reduction in the volume of blood ejected per beat. The reduction in stroke volume with rise in rate is shown in table 21. If, on the other hand, the venous inflow is large, the heart fills earlier in diastole and a period of diastasis (p 209) is created. Then any increase in rate by abolishing or reducing this period must increase the output (see also p 263).

Referring again to table 19, in which it is seen that the increased output is unaccompanied by a notable acceleration of the heart, it must be remembered that the experiments were carried out upon a denervated heart-lung preparation, the Bainbridge reflex (p 246) was in consequence abolished. It cannot be concluded therefore that in the intact animal the rise in venous pressure which resulted from the greater inflow would not have caused an increase in the rate of the heart. In the intact animal a moderate increase in the

TABLE 21

Dog 8.5 kg Weight of heart 77.5 gm Small constant venous inflow, heart rate increased The output remains unchanged

(Markwalder and Starling)

TIME	TEMPERATURE	RATE OF HEART PER MINUTE	ARTERIAL PRESSURE	SYSTEMIC OUTPUT	CORONARY SINUS OUTPUT	TOTAL CORONARY OUTPUT (CALCULATED)	TOTAL OUTPUT (CALCULATED)	OUTPUT PER HEART BEAT	VENOUS PRESSURE
	°C		mm Hg	cc per minute	cc per minute	cc	cc	cc	cm H ₂ O
2 20	28 2	72	119	612	24 20	40 3	652 3	9 05	11 0
2 28	30 2	90	119	612	20 62	34 4	646 4	7 20	9 0
2 37	33 4	114	119	612	18 46	30 8	642 8	5 63	6 2
2 42	35 0	126	119	625	19 74	32 9	657 9	5 2	5 0
2 50	37 0	144	117	612	21 20	35 3	647 3	4 47	4 2
2 58	39 0	156	116	612	22 90	38 1	650 1	4 16	3 8

output may also be brought about without an increase in the diastolic volume, the ventricles simply emptying more completely. In moderate exercise, in man, for example, the diastolic size of the heart as observed radioscopically may not increase, whereas the systolic size is less than that during rest.

The experiments of Starling and his colleagues have shown that the dog's heart is capable of pumping in the neighborhood of 57 times its own weight of blood per minute for a considerable length of time. In one instance a heart weighing 42 grams gave an output of 2400 cc per minute. If the power of the human heart may be taken to be equally great in proportion to its weight, then a heart weighing 300 grams or so should be capable of an output of over 17 liters per minute. The maximal output actually observed in man by indirect methods considerably exceeds this figure (p. 265). During an ordinary life-time the human heart performs an amount of work equivalent to that which would raise a weight of 100 tons to the height of 1 mile.

THE SOURCE OF THE ENERGY WHICH ENABLES THE HEART TO INCREASE ITS OUTPUT THE LAW OF THE HEART

It is well known that skeletal muscle contracts more forcibly if it is loaded by a weight before it is excited. The weight stretches the muscle fibers, i.e., it increases their length and exerts a certain tension upon them. If the muscle is made to contract isometrically the tension developed during the contraction is found to be proportional to the length of the muscle before excitation. The latter is called the *initial length* of the muscle. The tension which the load exerts upon the fibers

just prior to their contraction is spoken of as the *initial tension*. The tension developed when the muscle contracts—isometrically—will be referred to as the *developed tension* which, of course, is a measure of the force of the contraction. When a resting muscle is weighted, little change in initial tension actually occurs until it is extended beyond a length corresponding to that which it possesses when in its natural position in the body, i.e., its physiological length. Up to this point the developed tension increases with each increment in initial length, but beyond it the developed tension actually becomes less with increasing initial length. Yet, it is only when the muscle is stretched beyond its physiological length that any marked increase in initial tension occurs. The power of contraction of a skeletal muscle is therefore dependent upon initial length and not upon any stimulating effect exerted upon the muscle fibers by initial tension. The relationships between the initial length, initial tension and the developed tension of skeletal muscle during its contraction are shown in figure 25.5.

There has been some controversy as to whether the energy liberated by the cardiac muscle is, like skeletal muscle, dependent mainly upon initial length or upon initial tension. The experiments of Starling and his colleagues indicate that initial length is the sole determining factor.

Clearly, there are no means, as in the case of skeletal muscle, by which the length and tension of the cardiac muscle during its relaxation (diastole) or the tension developed during contraction (systole) can be measured *directly*. Yet, the diastolic volume depends upon fiber length (initial length), the intraventricular pressure during diastole represents initial tension, and the pressure

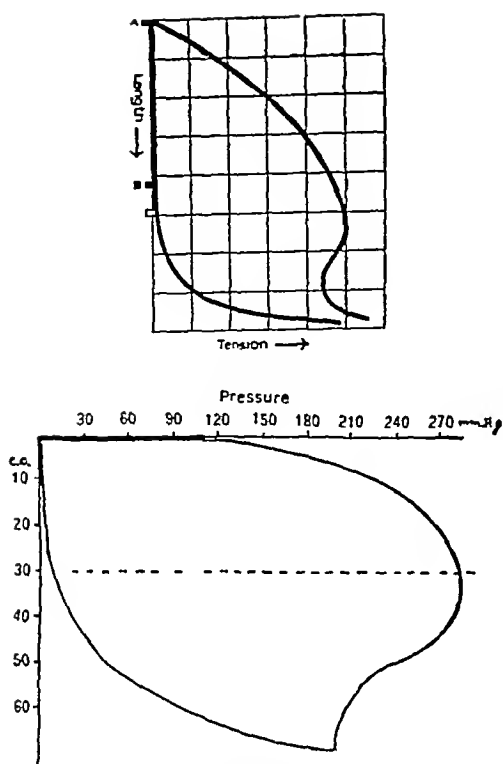


FIG 25.5 Showing (upper) the relation of initial length to the tension developed by skeletal muscle contracting isometrically, and (lower) the relation of diastolic ventricular volume (initial length) to systolic pressure (see text) (After Starling and associates)

developed during systole is related to the tension set up by the contracting fibers (developed tension). When, therefore, simultaneous records of diastolic volume and intraventricular pressures

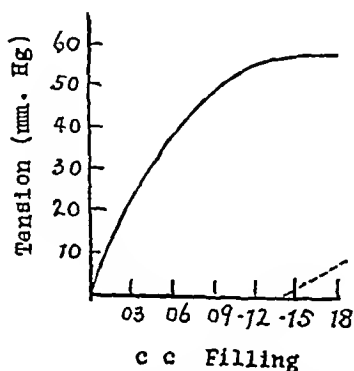


FIG 25.6 Showing the effect of graded increases in filling of the frog's ventricle upon the tension developed in isometric response, broken line, initial tension (Redrawn from Anrep and Segall)

(p 207) during diastole and systole are secured, the data may be plotted as shown in the diagram (fig 25.5) with volume (cc) representing initial length, along the ordinates and pressure (mm Hg) representing tension, along the abscissae. The venous pressure which stretches the cardiac muscle during diastole corresponds to the weight applied to the skeletal muscle. The opening of the semilunar valves against the aortic pressure represents the load against which the cardiac muscle has to contract. It will be seen from the figure that the force of the ventricular contraction increases with the diastolic volume until the heart becomes overdistended and a rise in diastolic pressure occurs. The force of the contraction then falls off (compare with fig 25.5), the two curves approaching one another and finally meeting.

Starling has expressed this fundamental principle of cardiac behavior in what has been termed the *Law of the Heart*, namely, that "the energy set free at each contraction of the heart is a simple function of the length of the fibers composing its muscular walls." Thus the heart fibers automatically gain the necessary energy to eject the greater load of blood which fills its cavities during diastole or to discharge its contents against a raised arterial pressure. In the latter instance the ventricle for the first beat or two after the pressure has been raised does not expel the full quantity of blood, but retains a part, this serves to increase the diastolic volume beyond that existing before the pressure was raised, and so to increase the force of the subsequent contraction. The greater energy liberation associated with the increased diastolic volume is attributed to the greater extent of chemically active surface which naturally results from elongation of the fibers.

The results of the experiments of Anrep and Segall are also confirmatory of Starling's conclusions. These observers found that when the isolated frog's ventricle contracted isometrically the contractile force (systolic pressure), increased up to a point, proportionately with the filling of the ventricle (diastolic volume), the maximal tension developed when the ventricle was filled to $\frac{2}{3}$ of its maximal capacity. Filling beyond this caused a rise in the initial (diastolic) tension accompanied by a reduction in the developed tension (fig 25.6). Katz was able to dissociate the effects of these two factors upon the development of tension by the turtle heart. When initial tension was varied but initial length kept constant, or the converse, or when both were varied in the same or in opposite directions, the results always indicated that initial length was the

factor which determined the force of the contraction. Changes in initial tension amounting to over 200 mm H₂O were usually without effect upon the height of the intraventricular pressure provided the diastolic volume was kept constant.

It has also been shown by Starling and Visscher that the oxygen consumption of the heart muscle, i.e., the total energy expenditure, is directly proportional to fiber length (diastolic volume). The ratio $\frac{\text{O}_2 \text{ consumption}}{\text{diastolic volume}}$ remained constant though

diastolic volume varied widely. The work performed by a heart in first-class physiological condition also bears a linear relationship to diastolic volume and consequently to the oxygen consumption. On the other hand, when as a result of fatigue the condition of the heart deteriorated, its diastolic volume was much greater in proportion to the work performed than was the case with the well-conditioned heart, i.e., the ill-nourished muscle fiber in order to gain energy for the performance of a given amount of work must be stretched to a greater extent. Nevertheless, whether the condition of the heart muscle was good or bad the relationship between oxygen consumption and diastolic volume was the same (fig 25.7). This means, clearly, that for the performance of a given piece of work the poorly nourished heart uses more oxygen (since it dilates more) than does a heart in good condition, or put in another way, the proportion of the total energy expenditure which appears as mechanical work (its efficiency) is lowered when the heart muscle departs from its prime physiological state.

PATHOLOGICAL PHYSIOLOGY OF CARDIAC LESIONS CARDIAC DILATATION AND HYPERTROPHY

The enlargement (dilatation and hypertrophy) of the diseased heart is a compensatory reaction. In aortic regurgitation, for example, the heart receives blood not only from the auricles but also from the aorta as a result of the incompetence of the aortic valves. The diastolic enlargement of the ventricular cavity which is required for the accommodation of the greater blood mass also enables the heart to develop the energy necessary for the ejection of the extra blood during systole. Cardiac dilatation therefore whether under physiological conditions (physiological dilatation) or in association with heart disease (pathological dilatation) is the means whereby the heart mobilizes

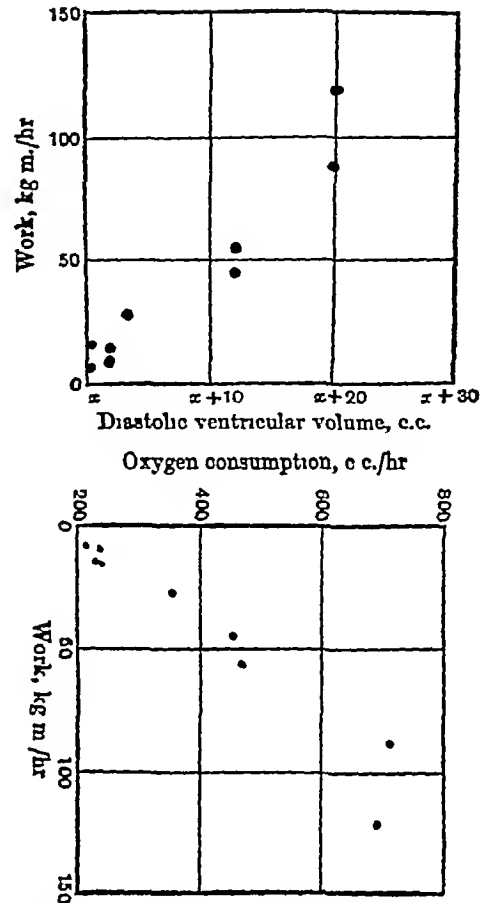


FIG 25.7 Lower chart shows the work done plotted against the oxygen consumed. In calculating the work of the heart the velocity factor (p. 142) was neglected. Upper chart gives data from the same experiment, work being plotted against the diastolic volume of the ventricles. The volumes are expressed as x plus known values. The x represents the lowest value of the volume during the experiment, which is impossible to measure when a cardiometer is used to record heart volume. The cardiometer enables one to measure only an increase over this minimum value. The figure shows a direct correspondence between work done and ventricular volume. (After Starling and Visscher.)

its reserves of energy (fig 25.8). The so-called reserve power of the human heart, i.e., its capacity for work, resides in the extensibility of its muscle fibers, within the physiological limit. It is apparent then that the nearer the fiber during diastole approaches its maximal physiological length the greater will be the encroachment upon the heart's reserve. A well-developed and efficient heart, in order to gain sufficient power to accomplish a certain amount of extra work, need dilate relatively little and is easily capable of discharging as much blood as it receives. The venous pressure in consequence shows little tendency to rise and the pulse is not greatly accelerated. On the other hand, the smaller heart or the heart with a myocardium

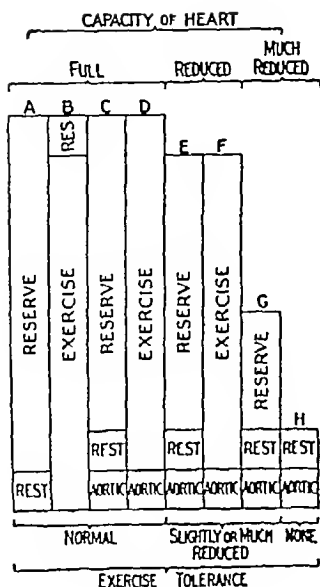


FIG. 258 Diagram illustrating the reserve of the normal heart and of the heart in aortic regurgitation (after Lewis). In column A the reserve power of the normal heart, i.e., its capacity for work, is represented as being several times greater than that required for rest. In column B the greater proportion of, but not all, the reserve is drawn upon to perform the extra work entailed in strenuous muscular exercise. Column C represents the heart in aortic regurgitation but with an efficient myocardium. It will be noted that the work required of the heart during rest is increased but its reserve is diminished to only a small extent, it is therefore capable of performing the extra work required for vigorous muscular exercise (column D). In columns E and F the reserve of the heart muscle is represented as falling short of the requirements of strenuous exertion. The reserve is reduced still further in column G, and in column H it has gone entirely, the heart muscle being capable only of meeting the requirements of rest. The subject is dyspneic during rest, and congestive heart failure is imminent or has already supervened.

weakened by disease, ill-nourishment or oxygen lack, when given an equivalent amount of work to perform must dilate to a greater extent in order to liberate the required energy. Indeed, an ill-equipped heart may, as a result of some extra burden, be dilated to its physiological limit and still be unable to increase its output per beat. The heart accelerates its beat as this point is approached (Bainbridge reflex) in order to increase its output. When the full physiological length of the fiber has been attained and the optimal pulse frequency developed, the heart has reached the limits of its powers. Blood then accumulates in the cardiac chambers, the venous pressure rises, the circulation through the capillaries is slowed

and the blood gives up a greater proportion of its oxygen load to the tissues (stagnant type of anoxia, p. 420).

THE PERMANENT EFFECTS OF EXERCISE UPON THE HEART

The belief has been widely held that strenuous muscular effort is conducive to cardiac dilatation and hypertrophy—signs of myocardial damage. A proportion of athletes have, of course, suffered from heart disease and an occasional racehorse has died with a dilated and hypertrophied heart. Instances of this sort have been cited in support of such a view. It is now agreed, however, that the cardiac enlargements under these circumstances are the result of preexisting disease and that the healthy heart cannot be dilated beyond its physiological limit, nor will a bout of strenuous exercise “strain” the healthy heart muscle or cause it to fail. In a healthy person the functional capacity of the heart and of the skeletal musculature are apparently so proportioned to one another that the greatest load of blood (venous inflow) which the muscles can provide is taken care of by the reserves of cardiac energy. The healthy heart of the marathon runner or oarsman does not hypertrophy to a pathological degree. It has also been shown by X-ray examination that the diastolic diameters of the healthy heart are not materially increased during strenuous exercise and that the systolic size is actually less than it is during rest, that is, the heart empties more completely. It may, therefore, be concluded that acute dilatation or chronic dilatation with hypertrophy is the result of some diseased state of the heart itself.

It should not be concluded, however, that muscular training exerts no effect upon the size of the heart. Mitchell, for example, found that the heart size of a group of undergraduates at Cambridge underwent a gradual increase over a period of a few years of athletic training. This was accompanied by a reduction in pulse rate (see p. 241). Lindhard observed an increase of 20 per cent in the resting cardiac output as a result of physical training. Others have reported more marked effects in persons indulging in very arduous types of athletics. Eyster, on the other hand, could demonstrate no increase in the average heart size of a group of young athletes when compared with that of a group following a more sedentary life. The evidence taken in review indicates that athletic pursuits, especially of a strenuous nature,

may cause a moderate but definite increase in cardiac bulk. This, however, is purely physiological and is proportioned to, or only slightly in excess of, a development of the skeletal muscles. That is to say, the normal ratio of heart weight to body weight shows no change or a very moderate one as a result of athletic training.

THE NATURE AND CAUSE OF THE HYPERTROPHY IN CARDIAC DISEASE

The average weight of the normal adult heart is 300 grams (250 to 350), it constitutes from 0.40 to 0.50 per cent of the body weight. In disease 500 grams is not an unusual weight and hearts weighing 1000 grams or more are occasionally seen. The commonest causes of cardiac hypertrophy are valvular disease and hypertension. In the latter the left ventricle is primarily involved but in valvular disease one or other ventricle, or both ventricles may be enlarged, in some instances the auricles are also dilated and their walls hypertrophied. Other less common causes of cardiac enlargement are coronary disease, arterio-venous aneurysm, hyperthyroidism, adherent pericardium, anemia and congenital cardiac defects. When the enlargement is due predominantly to hypertrophy it is sometimes referred to as *concentric*, when marked dilatation exists as well, the enlargement is said to be *eccentric*. The increased bulk of the ventricular muscle is due, not to a multiplication of its fibers, but to an increase in the size of each fiber. The sarcoplasm is more abundant and the fibrillae more numerous. There is also frequently an increase in the connective tissue between the fibers.

The hypertrophy is probably always preceded by some degree of dilatation. The permanent lengthening of the fibers consequent upon the latter, as already mentioned, tends to reduce the cardiac reserve since they are thereby brought nearer to the point of maximal extension. The hypertrophy, on the other hand, has the effect of increasing the power of the fiber at this new length with the result that the heart's reserve power is raised and may be restored nearly to its normal value (fig 25.8). Such a heart is said to be compensated. In experimental lesions the hypertrophied heart is reported to have a reserve power quite as great as that of the normal heart.

The *cause of the cardiac hypertrophy* is not easily explained. Increased work (induced by valvular defect, hypertension, etc.) is not the only, or

even perhaps the major, factor in its production. In the first place, the extra burden thrown upon the heart by a valvular lesion is often not sufficient to encroach to any great extent upon the reserve power of the heart, and in other instances the grade of hypertrophy does not correspond to the amount of extra work which the heart is called upon to perform. Arterial hypertension, if it develops gradually may, for example, cause little hypertrophy though the work of the heart is increased by 40 per cent. Also, in a large proportion of cases of mitral stenosis the hypertrophy of the left ventricle is as great as that of the right where the burden falls, and in aortic disease the right as well as the left ventricle may be increased in bulk (Lewis). Finally, cardiac hypertrophy occurs, as in coronary disease, when there is no evidence of increased work.

The experiments of Eyster and his associates throw some light upon this question. They produced aortic regurgitation in dogs by passing an instrument down the right carotid artery to the heart and cutting the aortic valves. The ventricle, being suddenly subjected during its relaxation phase to the high aortic pressure, became acutely dilated. The dilatation (as shown by roentgenograms) disappeared within a few days but hypertrophy gradually developed from then on. An effect comparable to aortic stenosis was also induced by constricting the ascending aorta with a rubber band. Similar changes in heart size resulted. In other animals the band was removed after from 3 to 6 days, at which time the ventricular dilatation had reached its maximum. Though the heart was relieved of the extra work, hypertrophy was not prevented. This came on gradually, as in those instances in which the stenosis was not relieved, and was complete by about the eightieth day. A second constriction again caused immediate dilatation followed by progressive hypertrophy. Similar results were obtained by massive transfusions which caused distention of the right ventricle. Histological examination of the heart during the stage of dilatation revealed thinning and hydropic degeneration of the heart muscle, which were ascribed to the injurious effect of the acute stretching of the muscle fibers. In the later stages the degenerative changes had disappeared and a pure hypertrophy alone was evident. The hypertrophy was therefore considered to be a reaction to injury brought about by the sudden dilatation.

These observations are of the utmost importance in gaining an insight into the probable sequence of events in the hypertrophy of the human heart. Overstretching of the fibers is apparently associated in some way with their overgrowth. The increased fiber length may be brought about by an excessive load resulting from some mechanical defect such as aortic regurgitation. Injury to the muscle is thereby induced. Or, on the other hand, the overstretching may be the result of a primary weakening of the myocardium, as in coronary disease or infective or toxic processes, the fiber being forced to lengthen unduly in order to liberate the required energy. The manner in which increase in fiber length brings about the overgrowth is unknown, but the experiments of Starling and Visscher suggest elevation of the metabolism of the muscle as a possible explanation (p. 255). It has been mentioned that, for the performance of a given amount of work, the ill-conditioned myocardium, as compared with a healthy one, has a much greater oxygen consumption. This view accords with the facts that a diseased myocardium may hypertrophy even though its work is not at all increased, and that a great increase in the burden of the healthy muscle can be borne with only a very moderate degree of overgrowth.

HEART FAILURE, CHRONIC CONGESTIVE HEART FAILURE (CARDIAC DECOMPENSATION)

The heart may fail suddenly, as under chloroform anesthesia, during an attack of paroxysmal tachycardia, as a result of occlusion of a coronary artery or one of its branches, following acute effusion of fluid into the pericardial sac, or from a sudden increase in cardiac work thrown upon an already diseased myocardium. In such instances there is a sharp reduction in cardiac output, the ventricles may enter into fibrillation and the circulation then comes to a standstill. Or, the patient may survive the attack which then may be followed by the symptoms characteristic of chronic heart failure.

CHRONIC CONGESTIVE HEART FAILURE

The heart with a valvular defect or one which is forced to work against some abnormal resistance may, if it possesses healthy muscle, continue to perform its functions efficiently for years. The state of the heart muscle rather than the valvular lesion itself has come to be recognized as the

important factor determining the onset of cardiac failure. The *healthy* myocardium possesses such great reserves of energy that there is little likelihood of its ever being faced with a resistance which it cannot effectively overcome—not until the myocardium itself becomes diseased do the signs of circulatory failure make their appearance.

Increased resistance to the output of the right ventricle results from such pulmonary conditions as emphysema or fibrosis of the lung, from pulmonary or mitral stenosis, or from failure of the left ventricle. Aortic stenosis, hypertension, or congenital narrowing (coarctation) of the aorta increases the resistance to the output of the left ventricle.

When a heart is persistently unable to maintain an adequate circulation during rest or mild exertion the condition is spoken of as chronic heart failure.

Causes of cardiac failure

Circulatory failure may be precipitated by the reduction, more or less suddenly, in the myocardial reserve of a heart which has been working against some inordinate resistance (e.g., hypertension), whose venous load is excessive (e.g., hyperthyroidism), or whose work is increased as a result of a mechanical defect (e.g., valvular disease). Conversely, failure may be induced by an increase in the burden of the heart whose reserve already has been lowered by myocardial disease. Indeed the heart muscle may be so weakened by disease—coronary sclerosis, acute infective processes, severe anemia, anoxia, etc.—that cardiac failure may supervene though no condition exists to cause an increase in the work of the heart. Among the commoner causes which may precipitate chronic cardiac failure are

(a) Infections, especially of the respiratory tract. The increased cardiac work resulting from the rise in metabolic rate caused by the fever, direct poisoning of the myocardium, and cough, all conspire to reduce the cardiac reserve to the point where failure ensues.

(b) Excessive muscular effort.

(c) Chronic pulmonary conditions, emphysema, bronchitis, etc., accompanied by much coughing which exerts its effect mainly through the muscular effort involved.

(d) Pregnancy (increased venous inflow).

(e) Rapid heart action, e.g., auricular fibrillation, paroxysmal tachycardia, etc. It should be remembered that the heart works very uneconomically at high rates of beating, i.e. the heart expends more energy for the performance of a given amount of work at high

than at low rates of beating. One reason for this is that the cardiac cycle is shortened, mainly at the expense of the diastolic (rest) period and the period of ejection (time during which work is actually performed). The period of isometric contraction (during which energy is expended in raising the intraventricular pressures above the arterial pressures but no work is done) is shortened comparatively little. This latter period is, therefore, in a sense a waste period, and its total time per minute is greater the more rapid the heart rate.

(f) Primary increase in blood volume (increase in venous inflow load). Even the healthy heart may fail if the venous inflow is greatly increased artificially as by a massive and rapidly administered transfusion.

The validity of the application to clinical cases of Starling's law of the heart, which was based upon acute animal experiments, has been tested by Starr and his associates. The work of the heart, as determined in 140 normal persons and in a number of patients who had recovered from congestive heart failure, was plotted against the heart volume. The work of the heart was calculated from the cardiac output and the blood pressure, the heart size was determined roentgenologically. In the cardiac patients, the work performed for a given heart volume was much less than in normal subjects. These authors concluded that Starling's law holds true for the human heart in congestive failure.

There is some difference of opinion concerning the fundamental cause of chronic heart failure. According to one view (Katz and associates), the total energy liberated at the maximal physiological length of the myocardial fiber is reduced to the point at which an adequate circulation cannot be maintained. Visscher, on the other hand, considers that, in the majority of cases at any rate, the heart fails because of a decrease in its mechanical efficiency (that is, because of a reduction in the proportion of the total energy output which is capable of being utilized in performing work). Starling and Visscher found that this was invariably the cause of cardiac failure in the heart-lung preparation of the dog. In one of their experiments of 4 hours duration, the oxygen consumption rose from 360 cc to 590 cc per hour, the diastolic volume increased by 42 cc and the efficiency fell from 9.5 per cent to 5.8 per cent.

THE SIGNS AND SYMPTOMS OF CONGESTIVE HEART FAILURE

The chief manifestations are increased blood volume (unless failure develops rapidly), raised

venous pressure, dyspnea, cyanosis, congestion and consequent enlargement of the liver and spleen, pulsation of the liver is not uncommon, and there may be jaundice, oliguria and albuminuria, and edema. The cardiac output is, as a rule, less than normal and responds in less degree than in health to a rise in the venous inflow (e.g., as induced by muscular exertion). This is especially evident in the more advanced stages of chronic failure, the cardiac output showing a very small rise, or may actually decline if the venous load is much increased.

The mechanism by which the chief signs and symptoms of chronic congestive heart failure are produced has been a question of considerable discussion. According to one view—known as the *forward failure theory*—they are primarily the result of a reduced cardiac output and the consequent inadequate blood supply to the tissues. The dyspnea, for example, has been attributed to slowing of the circulation through the medulla (respiratory center), the cyanosis to the diminished flow through the vessels of the skin and the greater coefficient of oxygen utilization. Diminished renal blood flow (reduced glomerular filtration) leads to the retention of salt and water, increased blood volume (anoxia of the bone marrow also stimulates the output of red cells), edema, and a rise in venous pressure (see also p. 37). Subnormal blood flow to the skeletal muscles causes weakness and easy fatigability. In acute cardiac failure, as from coronary thrombosis or other cause, there is a marked reduction in cardiac output. Venous congestion and edema are not seen, and the manifestations—pulsus alternans, gallop rhythm and cardiac asthma—are readily explained upon the forward failure theory—left ventricular failure.

In chronic congestive heart failure, however, the cardiac output is not always reduced, and when associated with certain diseases, e.g., severe anemia, hyperthyroidism or pulmonary emphysema, it may be actually increased. Again in many cases of cardiac failure, when a reduced cardiac output does occur, the reduction does not seem sufficiently great to account for the symptoms on the basis of "forward failure." In most patients with congestive heart failure the clinical picture, according to Harrison, is interpreted best upon the "*backward failure*" theory (back pressure theory), which now may be briefly stated.

As the myocardium fails, blood accumulates in the ventricle which in consequence dilates. The

muscle fibers become stretched to a point where they offer a greater resistance to the incoming venous blood, a rise in auricular pressure follows which is transmitted backwards. The heart does not fail as a whole, failure of the left ventricle may precede failure of the right by a considerable interval or death may occur from failure of the left or the right ventricle alone.

When the muscle of the left ventricle is the first to weaken, a greater amount of blood than normally remains in the ventricle at the end of systole. This residual blood together with that received during and following diastole stretches the muscle fibers. The pressure of blood in the left auricle and pulmonary veins rises. The high pressure in the pulmonary circuit causes engorgement of the vessels of the lung. The velocity of flow in the individual vessels is diminished. The distended vessels encroach upon the air spaces, the vital capacity is reduced, the lung tissue is rendered less expansile, dyspnea and cough result (p 416). Increased resistance in the pulmonary circuit, such as results from stenosis of the mitral orifice, even in the absence of left ventricular failure tends also to cause pulmonary engorgement accompanied by dyspnea, especially upon exertion. A much larger than normal proportion of the circulating blood volume is held in the lungs. This fact and the slowing of the renal blood flow with the retention of water and salt leads to a secondary compensatory increase in the total blood volume. Stimulation of afferent nerve endings in the great veins at the base of the heart by the high venous pressure is also a probable factor in the production of dyspnea.

So long as the right ventricle continues to contract forcibly and discharge its contents against the increased resistance in the pulmonary circuit, the systemic vessels are not congested. Subsequent failure and dilatation of the right ventricle are attended by a rise of right auricular pressure and of the pressure in the systemic veins, with congestion of the liver and abdominal viscera. Signs and symptoms (edema, cyanosis, etc.) appear, due to engorgement of the peripheral vessels and slowing of the blood flow through them. But, though the velocity of the blood in the peripheral capillaries is reduced, owing to their distension and the consequent general enlargement of the capillary bed, the total volume of blood flowing through the periphery may show little change (see p 178).

The following observations are advanced against the forward-failure hypothesis as an explanation of the manifestations of the usual type of cardiac failure, and argue for the "back-pressure" theory.

(1) Reduction in the cardiac output, though commonly observed, has not been demonstrated invariably in congestive heart failure, the reduction in cardiac output, when a reduction does occur, does not run parallel with the severity of the symptoms. Moreover, the improvement of the clinical condition which follows the administration of digitalis is not always attended by an increased circulation rate.

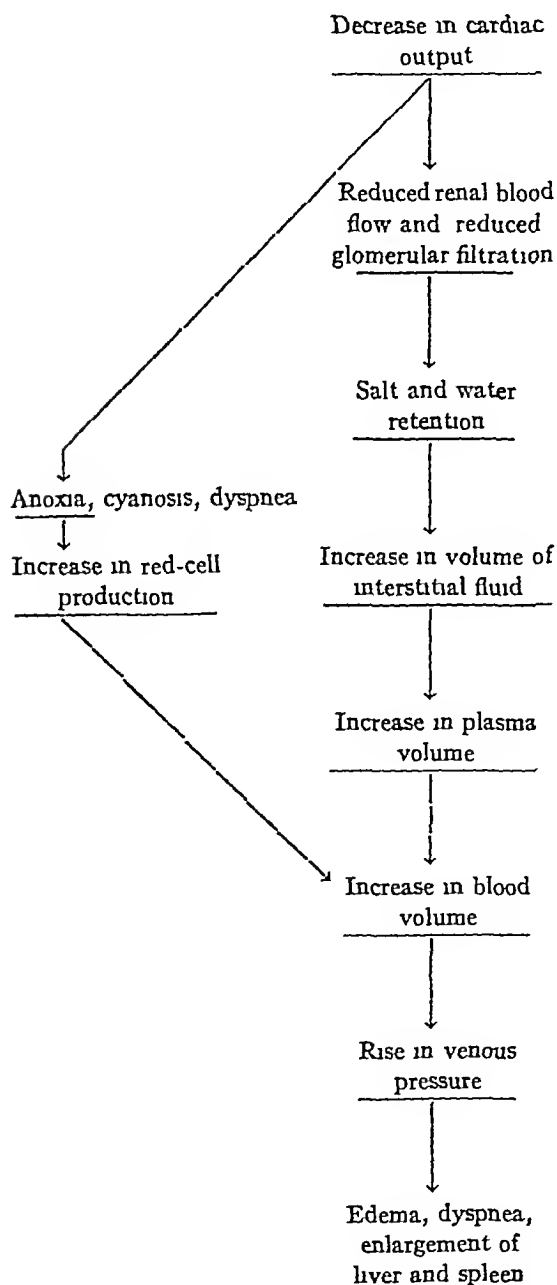
(2) The fact that one ventricle can fail independently of the other and that left-sided failure is associated with pulmonary congestion and right sided failure with systemic congestion can be explained upon the basis of "back-pressure", but not by the forward-failure hypothesis. The reduction in vital capacity in failure of the left ventricle or in obstruction (stenosis) at the mitral orifice is accounted for most satisfactorily by a rise of pressure in the pulmonary circuit and congestion of the lungs. Congestion of the lungs and diminished expansibility, as a result of compression of the pulmonary veins, have been demonstrated in animals.

(3) Reduction of the cardiac output in animals does not reproduce the ordinary picture of congestive heart failure, raised venous pressure, edema and certain other manifestations are absent.

Though a "backward failure" mechanism can account for the clinical picture in the majority of patients with congestive heart failure, it appears from the investigations of Stead and Warren, Merrill and others that "forward failure" explains the train of symptoms in a certain proportion of cases. The high venous pressure and edema in congestive heart failure is attributed not "to raising a dam" at the right ventricle and thus distending the systemic veins, but to a train of events which starts with a reduced cardiac output leading to reduced renal blood flow, salt and water retention and increase in interstitial fluid, and ends with enlargement of plasma (and blood) volume (p 38). It is pointed out that in congestive failure an increase in weight, due to water retention ("subclinical edema"), occurs before the rise in venous pressure. Furthermore, hemodilution is found coexistent with the edema, whereas if a primary rise in venous pressure were the cause of the edema, hemoconcentration (due to leakage of fluid from the circulation) would result. In a study of a series of patients with congestive heart failure without renal disease, Merrill found the cardiac output (determined by direct Fick method (p

268)) reduced to about one-half and the renal blood flow to one-fifth. The greater reduction in the renal blood flow as compared with that of the cardiac output suggests a shunting of blood away from the kidney. The reduction in the blood flow through the kidney was found to bear no relation to the venous pressure. Starr has also shown in experiments upon dogs that the most extensive damage of the right ventricle consistent with survival did not cause a rise in venous pressure.

The following scheme illustrates the "forward failure" theory.



Comparing the two theories, it will be seen that they differ mainly as to the origin of the raised venous pressure, the backward failure theory explaining it simply as a back pressure effect, whereas, according to the forward failure theory, it is secondary to an increase in blood volume caused in turn by a reduced renal blood flow, and the retention of water and salt. But it seems likely as already intimated that the rise in venous pressure is not always due to the same mechanism, one or other being operative in different cases.

The possibility must also be considered that tissue anoxia induces in some way an endocrine disturbance which is responsible for the retention of salt and water. It will be recalled that increased pressure in the renal vein, as would result from back pressure, causes salt retention in some obscure manner. The whole question is one of the utmost complexity.

The inefficiency of the heart in congestive failure
For the performance of a given amount of work the oxygen usage of the subject of congestive heart failure is considerably higher than that of a healthy person. From experiments upon animals (p. 255) one would suspect that the heart is responsible to some extent at least for the higher oxygen consumption. Resnik and Friedeman found that the basal metabolism of patients with heart failure was from 10 to 60 per cent above normal, and that improvement in the cardiac condition was accompanied by a reduction in the amount of oxygen consumed. They conclude that increased respiratory effort (i.e., the increased work of the respiratory muscles) accounts for the greater part, but that a considerable share of the extra oxygen is used by the heart itself. Other factors which have been suggested as being responsible, e.g., thyroid stimulation as a result of circulatory alterations in the gland, apprehension and discomfort, or expansion of the capillary bed caused by the high venous pressure, played no rôle in the opinion of these authors in producing the high metabolic rate.

Exercise tolerance tests A useful test of cardiac function in heart disease based upon the metabolic response to exercise has been devised by Katz and his associates. The greater proportion of the increased oxygen consumption occurs after the exercise. These observers, therefore, determine the oxygen consumption before, during and for 15 minutes after the performance of a measured amount of work. The total oxygen consumption throughout the test is compared with that for

the same period at the pre-exercise rate. The difference between the two estimations gives the excess metabolism (ch 52) of the exercise, which is expressed in cubic centimeters of oxygen consumed per kilogram-meter of work per square meter of body surface. A simple but valuable aid in estimating cardiac reserve in chronic heart disease is the determination of respiratory and cardiac acceleration after a standard light exercise, e.g., 50 hops 3 inches from the floor at one second intervals. In testing physical fitness in

apparently normal persons, e.g., of army recruits or applicants for insurance, similar but more exacting procedures are employed, such as stepping alternately off and on a stool or bench from 9-20 inches high. In healthy persons the heart rate taken immediately after a 30-second period of such exercise (12 inch bench) increases by from 20 to 30 beats per minute, but returns to normal within a minute of cessation of the work.

CHAPTER 26

THE OUTPUT OF THE HEART

DEFINITIONS AND GENERAL CONSIDERATIONS

The output of the heart per beat is spoken of as the *systolic discharge* or the *stroke volume*, and the output per minute as the *minute volume* or the *circulation rate*. The value of the latter is simply the product of the stroke volume and the pulse rate (pulse rate \times stroke volume = minute volume), the minute volume divided by the pulse rate, therefore, gives the stroke volume. The quantity of blood ejected by each beat of the left ventricle in the average healthy man during rest is from 70 to 80 cc. An equal quantity is, of course, discharged at the same time by the right ventricle, making a total for the whole heart of from 140 to 160 cc. The contents of the left ventricle are ejected against a much higher mean arterial pressure than the contents of the right, the mean pressure in the pulmonary artery being about $\frac{1}{3}$ of that in the aorta. The minute volume is expressed in terms of one ventricle. The output of one ventricle obviously represents the quantity of blood flowing through the lungs, or through the systemic vessels, during the same period. It is for this reason that the value expressing the minute volume of the ventricle is also referred to as the circulation rate. An adult of sedentary occupation pumps at least 5500 liters of blood through his body daily—from left to right ventricle through the systemic vessels, and from right to left through the lungs.

The heart cavities are not believed to completely empty themselves during the resting state, but contain around 100 cc of residual blood at the end of systole. Each ventricle has a normal capacity of approximately 200 cubic centimeters.

It is evident that the heart can increase its output per minute by increasing its stroke volume and maintaining its rate constant or, conversely, by increasing its rate and maintaining a constant stroke volume. Again, both these factors may be called into play. When the heart rate accelerates the output per beat can be kept constant only if the venous inflow is adequate, only under such circumstances can the minute volume be increased. If, for example, the beats of the heart have occurred at the end of a period of rapid filling, i.e., before or at the moment that the heart chambers are filled, simply increasing the heart rate will

cause the beats to fall earlier in the period of rapid filling. Reduction in the stroke volume in proportion to the increase in rate must result and no increase in the minute volume can, therefore, occur (see table 21, p. 253). At very rapid heart rates, as in paroxysmal tachycardia, a point is reached at which the heart does not relax sufficiently between beats to take on an adequate load of blood and the minute volume is actually reduced.

If, on the other hand, the beats have been occurring some time after the period of rapid filling, i.e., a period of diastasis exists, then an increase in heart rate will, by shortening or abolishing this period, or by preventing its appearance when the inflow is augmented, increase the output. Consequently, when the venous return is increased as in muscular exercise, the greater mass of blood is readily accommodated by the large well-developed heart and there may be no period of diastasis, no rise in venous pressure, the heart rate then shows less tendency to increase. The hearts of some athletes show no increase in rate, whatever, during exercise, the resting stroke volume of such hearts may be double or more that of an ordinary person, namely 120 to 130 cc. In the smaller heart the greater inflow causes the heart to fill earlier in diastole, the venous pressure rises and as a result of the Bainbridge reflex the heart accelerates. Little change in stroke volume may occur, the greater minute volume being brought about almost entirely by an increase in the frequency of the beats. In the majority of healthy persons both factors—increase in heart rate and increase in stroke volume—play a part in the production of a greater cardiac output. Individuals differ, however, in the extent to which each factor contributes. In the athlete, as just mentioned, the output is increased chiefly through an increase in the stroke volume, increase in heart rate plays a very minor rôle, whereas in the sedentary individual the latter factor exerts an important effect.

The contribution which cardiac acceleration can make toward the minute volume is strictly limited, for from 180 to 200 is about the maximum rate to which the healthy heart can be speeded up. This is only about $2\frac{1}{2}$ to 3 times the normal rate.

With a constant stroke volume, therefore, cardiac acceleration could not increase the minute volume more than two and a half to three times. In muscular exercise, the oxygen consumption is increased many fold. There are only two possible means whereby the tissues can be supplied with the extra oxygen, namely, by an increased circulation rate or a greater coefficient of oxygen utilization (i.e., the removal of more oxygen from each unit of blood, p. 374). Were the cardiac output solely dependent upon a rise in pulse rate, then the circulatory factor could, obviously, increase the supply of oxygen to the tissues three times, at the most, over that of the resting state. If no change in the stroke volume occurred, then, in order to increase the oxygen supply, say 10 times, the coefficient of oxygen utilization would require to be increased between three and fourfold. As a matter of fact, the cardiac output in a robust subject may increase nine-fold during strenuous exercise, and the oxygen consumption 12 times. In such an instance it must be concluded (1) that the stroke volume increases over 3 times (even at a maximum pulse rate), and (2) that the oxygen requirement is satisfied largely by a rise in the circulation rate, an increase in the coefficient of oxygen utilization playing a less prominent rôle.

The proportion in which each of the two factors just mentioned—increased circulation rate and increased coefficient of oxygen utilization—contribute in bringing about the greater oxygen supply to the tissues varies, however, in different persons and with the type of exercise. In very light exercise involving slow movements the coefficient of oxygen utilization is increased to a proportionately greater extent than is the cardiac output. With moderately severe exercise, on the other hand, the coefficient of oxygen utilization is usually about doubled while the cardiac output is increased fourfold or so. This would permit an eightfold increase in oxygen consumption. As the exercise becomes more severe the circulation rate increases to a proportionately less extent than the coefficient of oxygen utilization, and in exhausting effort a point is reached where the cardiac output shows no further increase, the coefficient of oxygen utilization from then on rising with the oxygen consumption. The rise in the coefficient of oxygen usage in muscular exercise is attributed to the diversion of a larger proportion of the total blood volume through the contracting muscles. The production of acid metabolites (carbon dioxide and lactic acid) and

the rise in temperature also lower the affinity of the hemoglobin for oxygen which, in consequence, is given up more readily (p. 375) to the tissues.

The extent to which these two factors contribute to satisfy the oxygen requirement also varies with the particular muscle groups which are exercised. Muscles, such as those employed in walking or running, whose action is more efficient in increasing the venous return, tend to increase the cardiac output more than those which have not this effect, even though the total oxygen consumption in both instances is the same. Swimming is an exercise which causes a relatively great increase in the cardiac output. Several factors other than the exercise itself, namely, increased heat loss, respiratory stimulation and the pressure of the water are largely responsible for the greater cardiac response in this type of exercise. There is some evidence that, as a result of training, the coefficient of oxygen utilization tends to increase during exercise, thus sparing the work of the heart.

THE OUTPUT OF THE HEALTHY HEART NORMAL STANDARDS PHYSIOLOGICAL VARIATIONS

The minute volume of the heart under basal conditions¹ varies, in different individuals, from 3 to 4.6 liters. The value for a given person was found by Grollman to be remarkably constant, determinations made from time to time varying by no more than 2.5 per cent. Grollman has also shown that the basal cardiac output is a function of the surface area of the body. For normal persons the minute volume per square meter of body surface—the *cardiac index* as it is termed—has an average value of 2.2 liters (acetylene method p. 269). In a group of 50 normal adults the index had an average deviation from the mean of 6.4 per cent. The cardiac output then is proportional to the basal metabolism (p. 617) and like the latter can be predicted for a normal person, with a small error, from the surface area. The average basal cardiac output per kilogram of body weight is 62 cc. As already mentioned the average volume of blood ejected per beat (stroke volume) is from 60 to 70 cc.

The value of the cardiac index as determined by the direct Fick method is about 27 per cent higher than that found by the acetylene method.

¹ I.e., with the body recumbent and at rest, at room temperature 20°C and 12 hours after having partaken of food or drink.

Cournand obtained an average figure for the cardiac index by the former method of 3.12 liters, which gives an average total minute volume for adults, under basal conditions, of approximately 5.5 liters. The cardiac index is somewhat higher for children.

The following physiological conditions vary the cardiac output

- ① Muscular exercise In strenuous exercise the output increases from the basal level of between 5 and 6 liters per minute to between 19 and 37 liters, according to the individual, the stroke volume increases from a resting value of from 60 to 70 cc. to from 100 to 200 cc. The effect of muscular exercise upon the circulation rate is considered in more detail above.
- ② Temperature No change in cardiac output occurs until the environmental temperature exceeds 30°C. Temperatures above this cause a very moderate increase (5 to 30 per cent) in minute volume.
- ③ Digestion of food causes an increase above the basal level of from 30 to 40 per cent. This level is reached about 1 hour after the meal, persists for about 3 hours and then gradually declines. The extra strain placed upon the diseased heart following a meal probably accounts for the attacks of angina pectoris (p. 328) which sometimes occur at this time. Ingestion of fluids also increases the cardiac output to a moderate degree. Both the cutaneous and splanchnic regions share in the increased blood flow.
- ④ Extreme variations in oxygen and carbon dioxide tensions in the inspired air (pp. 250 and 249).
- ⑤ Sleep causes little (10 per cent) or no reduction in the minute volume below the basal level.
- ⑥ Posture Schneider and Crampton find that the minute volume is greater in the recumbent than in the erect position with the subject standing quietly. After prolonged standing, in most healthy persons the output of the heart remains unchanged or decreases slightly. In certain subjects, who upon prolonged standing show poor circulatory compensation for the effect of gravity, the pulse pressure decreases by 20 mm Hg or so and the cardiac output is reduced markedly, fainting may occur. In congestive heart failure this increase in minute volume on changing from the upright to the recumbent position does not occur (McMichael).
- ⑦ Anxiety, anger and, probably, certain other emotional upsets. The effect of anxiety or apprehension is important to bear in mind in deter-

minations of the cardiac output in nervous subjects. The increased output in these cases may be due to the liberation of adrenaline, and may possibly occur without an increase in pulse rate, for McMichael and Sharpey-Schafer observed that an increase in output, amounting to 50 per cent, was induced in human subjects by the continuous infusion of physiological amounts of adrenaline, though there was no rise in heart rate.

⑧ Pregnancy increases the minute volume, but only in the later months is the increase pronounced, the maximum increase amounts to from 45 to 85 per cent. The increase is attributed to the large blood volume in the pregnant state and, as a consequence, to the greater venous inflow load.

PATHOLOGICAL VARIATIONS IN THE CARDIAC OUTPUT

A Conditions which increase the cardiac output

(1) Hyperthyroidism. It has been mentioned that the basal cardiac output is proportional to the basal metabolic rate. It therefore follows that the former will be increased in hyperthyroidism, as a matter of fact the output in this condition is raised from 50 to 100 per cent. The shunting of blood from the arterial to the venous side through the dilated thyroid vessels may possibly be a contributory factor (ch. 58) in the production of the increased output. After operation the output falls with the decline in the metabolic rate. Within the last few years thyroidectomy has been employed in angina pectoris and congestive heart failure with the object of lowering the metabolic rate and so of reducing the work of the heart. This mode of treatment is resorted to even though no evidence of hyperthyroidism exists. The thyroid is removed completely, thyroid extract is then administered with the aim of maintaining the basal metabolic rate at a level a little above that at which signs of hypothyroidism appear (-25 to -30). The operation is of benefit in angina pectoris, but its justifiability in other cardiac conditions is seriously questioned. Thiouracil (chap. 58) has been employed for the same purpose.

(2) Anemia (see p. 420) increases the output.

(3) Anoxemia increases the cardiac output unless severe or prolonged, then the output declines owing to the injurious effect upon the myocardium.

(4) Fever. The increase is due mainly to the elevated metabolism (ch. 53).

(5) Angina pectoris During the attacks of pain, the cardiac output is definitely increased over that in the periods between attacks

(6) Arterio-venous fistula (arterio-venous aneurysm) A communication between a large vein e.g., the femoral, and its companion artery results in a fraction ($\frac{1}{3}$ – $\frac{1}{4}$) of the blood ejected from the left ventricle being short-circuited, or shunted, to the right side. The velocity of the blood on the proximal side of the fistula is increased, that on the distal side reduced. Trauma is the most common cause of the condition. A pulsating swelling is frequently present at the site of the anastomosis, the heart rate is elevated, the heart becomes enlarged and, according to Harrison and associates who studied the condition experimentally, the minute volume is increased. An increase of 100 per cent in the cardiac output was found in dogs in which a femoral arterio-venous anastomosis had been produced. Compression of the artery or closure of the stoma causes an immediate reduction in the heart rate. Grollman from studies of the condition in man also concluded that the cardiac output was increased.² The effects upon the arterial system of an arterio-venous anastomosis resemble those of aortic regurgitation—low diastolic pressure, high pulse pressure (100 mm) and collapsing pulse—though the blood in the one case leaks into the ventricle during diastole, in the other into the venous system. Capillary pulsation occurs in both conditions.

The acceleration of the pulse is ascribed by Lewis and Drury to the reduction in the mean arterial pressure (Marey's law, p. 245) resulting from the leak through the arterio-venous fistula, rather than to an increase in venous pressure (Bainbridge reflex). They found that after vagal tone had been reduced or abolished by atropine, compression of the artery caused a reduction in rate of only two beats per minute. This indicates that the reduced rate, which occurs when the artery is compressed without previous treatment with atropine, is due to a vagal reflex. The enlargement of the heart is considered to be due to deficient nourishment of the myocardium caused by the lowered aortic pressure and the consequent reduction in coronary blood flow (p. 328).

(7) Page's disease of bone A large increase (to over 13 liters) in minute volume occurs in this

² Ellis and Weiss (Am Heart J 1930, 5, 3) observed in patients only a slight tendency towards reduction in the cardiac output when the aneurysm was compressed or after operative cure.

disease, due to augmented blood flow through the bones (Edholm and associates)

(8) Beri beri

B Conditions which reduce the cardiac output

(1) Cardiac irregularities (a) Paroxysmal tachycardia Biercroft and his associates found the output reduced to half the normal value in a subject of this condition. Before the onset of the attack the pulse rate was 64 per minute, the stroke volume 77.5 and the minute volume about 5 liters. During the attack the pulse rate was 198 and the stroke volume 12.9 cc., the output per minute was therefore 2.5 liters. (b) In auricular fibrillation the output is frequently reduced since this irregularity so often accompanies myocardial failure (see below). (c) In complete heart block with an efficient myocardium the output under ordinary circumstances shows little alteration from the normal. Since the pulse rate is from 35 to 40 per minute the stroke volume is elevated well above the normal average (100 cc. or more). If the stroke volume does not increase sufficiently to compensate for the slower heart rate, reduction in the minute volume is of course, bound to occur. In other instances associated with coronary sclerosis and myocardial degeneration, the output is definitely reduced. In other subjects again, though the output during rest is normal, the inability of the heart to accelerate fully prevents an adequate output during muscular exertion.

(2) Valvular disease and myocardial failure. In the absence of cardiac failure, a valvular lesion, as a rule, causes little change in the output under basal conditions, a reduction is, however, more likely to be found in mitral stenosis than in other forms of valvular disease. Cases which show no reduction in output during rest may, nevertheless, show a much smaller increase in minute volume in response to exercise than does the normal person, the tissues consequently suffer from oxygen lack and cyanosis may appear. The output also remains above the resting level for some time after the exercise has ceased, whereas normally it falls quickly to its previous value. In other subjects of valvular disease but with an efficient myocardium the cardiac output during exercise is not below normal, but, after the termination of the effort, it usually, as in the previous group, returns more slowly than is normal to the resting level. When definite weakening of the myocardium supervenes and other signs of heart failure are present, the

basal cardiac output is usually reduced, in some instances, however, little or no reduction in output occurs (see p 260) The increase in cardiac output, which normally results from muscular effort, does not occur or is slight in patients with congestive heart failure, but this failure is to some extent compensated for by a greater increase in the arterio-venous oxygen difference than is seen in healthy persons undergoing the same grade of work

Overweight persons with cardiac failure should be benefited by a reduction in their weight. It has been found by Master and his associates, in a series of persons not suffering from heart disease that a reduction of 16 per cent in oxygen consumption, a 30 per cent decrease in the cardiac output and an average reduction of 35 per cent in the work of the heart may follow loss of weight

(3) Myxedema The reduction in the cardiac output is roughly proportional to the depression of the basal metabolic rate

(4) Adherent pericardium and pericarditis with effusion The reduced minute volume is the result of the mechanical interference with the action of the heart. Adherent pericardium interferes with ventricular systole. Fluid within the pericardial sac prevents full dilation of the heart during diastole and thus reduces the volume of blood which it can accommodate

(5) Pneumothorax and other pulmonary conditions in which the intrathoracic negative (suction) pressure is reduced (p 357)

(6) Hemorrhage and surgical shock. Reduction in the cardiac output occurs early (p 301) in shock

(7) Arterial hypertension In some cases the minute volume is reduced, in others, probably the majority, it is within the normal range

(8) Postoperative A pronounced reduction in the cardiac output occurs for a period of from 1 to 4 days following surgical operation (Snyder)

THE EFFECTS OF CERTAIN DRUGS UPON THE MINUTE VOLUME

Adrenaline and histamine increase the oxygen consumption and the cardiac output, the effect upon the minute volume is, however, proportionately greater than that upon the oxygen consumption. Acetylcholine, whose effects in general are very evanescent, causes a slight increase in the minute volume. A decided rise in the cardiac output is produced by nitrites which, like acetylcholine, cause arteriolar dilatation. The increase in the cardiac output caused by nitrites is probably a compensatory effect of the lowered pe-

npheral resistance, whereby the blood pressure is maintained near the normal level. Digitalis produces no immediate effect upon the minute volume in normal persons, but a slight increase occurs about 6 hours after administration of the drug, followed by a reduction which persists for twenty-four hours In subjects of congestive heart failure with auricular fibrillation the output is as a rule though not invariably increased as the cardiac condition improves. Strophanthin acts similarly to digitalis. Atropine which increases the heart rate to 150 or 180 beats per minute does not increase, as a rule, the cardiac output. Alcohol in moderate dosage (35 cc) causes no more than a slight rise in the cardiac output. Grollman believes the effect to be largely psychic in origin rather than due to the direct action of the drug upon the heart. Caffeine increases the cardiac output, pituitrin and morphine reduce it slightly

MEASUREMENT OF THE OUTPUT OF THE HUMAN HEART

Any method for the estimation of the circulation rate in man must, of course, be indirect. Several methods have been devised for the purpose. The methods fall into three groups, (1) those based solely upon the Fick principle, (2) those which involve breathing a foreign gas as well, e g, nitrous oxide, ethyl iodide or acetylene, and (3) physical methods

THE FICK PRINCIPLE

The output can be calculated from the difference between the oxygen (or CO₂) content of the venous blood and that of the arterial blood, and the total oxygen consumption (or CO₂ elimination). It must be evident that if the quantity of oxygen which a unit of blood delivers to the tissues (or takes up from the lungs) is known, together with the total quantity of oxygen consumed over a given period, then the volume of blood which had been engaged in the carriage of this quantity of gas can be calculated. To take an example. The arterial blood contains about 19 volumes of oxygen per 100 cc. It gives up, let us say, 6 volumes to the tissues, i.e., the mixed venous blood coming to the lungs contains 13 volumes per cent. The arterio-venous (A-V) oxygen difference is, therefore, 6 volumes per cent. If the total quantity of oxygen consumed per minute is found to be 250 cc, then the cardiac output is—

$$\frac{\text{(Total O}_2 \text{ consumption per minute)}}{\text{19 - 13}} \times 100 = \text{(Output per minute)} = 4.16 \text{ liters,}$$

(Arterio-venous O₂ difference)

The cardiac output can be calculated in a similar way from the total carbon dioxide elimination and the arterio-venous carbon dioxide difference (CO₂ in mixed venous blood less CO₂ in arterial blood). The total oxygen consumption (or carbon dioxide elimination) can be readily determined (ch 45). The arterial oxygen or carbon dioxide content can be ascertained from analyses of samples of venous and arterial bloods. This is the *direct Fick* method and, until recent years, has been applicable only to animals, venous blood being obtained by puncture of the right side of the heart, or by the passage of a catheter into the auricle through the jugular vein. Oxygenated blood was obtained by arterial puncture. But the gas contents of the arterial and of the venous blood can be determined *indirectly* from their tensions, which can be obtained in turn from an analysis of alveolar air. This is a fairly simple matter in the case of arterial blood, but it is difficult to determine the gas tensions of the mixed venous blood, i.e., of the blood reaching the lungs. It is not permissible simply to take the gas content in the blood of an arm vein, since this is not the same as that of the mixed venous blood in the right auricle. The amount of oxygen given up by the blood to different tissues varies widely, but the right chambers of the heart receive mixed blood from all organs and tissues, including the heart muscle itself.

The direct Fick method

The direct Fick method was first employed in man in Germany in 1929, by Forssmann who experimented upon himself, and a little later by Klein. It has been developed in America chiefly by Cournand and his associates. A sample of mixed venous blood is drawn from the right auricle, right ventricle, or from the pulmonary artery, by means of a ureteral catheter (No 8 or 9) passed up the median basilic, axillary and subclavian veins. The introduction of the catheter is guided by X ray visualization (fluoroscope). A sample of arterial blood is obtained by arterial puncture. The blood samples are analyzed for their oxygen contents and, after the total oxygen consumption has been determined in the usual

way (ch 45), calculation of the cardiac output is made from the formula already given. The carbon dioxide A-V differences may also be determined to serve as a check on the result obtained from the oxygen A-V differences. This method has a high degree of accuracy and has proved of the utmost value in the diagnosis of congenital cardiac defects (see p 433). The chief source of error is the possible failure, owing to streamlines of blood differing in their gas content, to obtain a sample of perfectly mixed venous blood. However, Cournand found that, except in a small percentage of determinations, samples taken from auricle and ventricle did not differ in O₂ content by more than 0.25 volume per cent. In competent hands the method appears to carry little serious risk. The cardiac index (p 264), as determined by the direct Fick method, is about 27 per cent higher than that obtained by the acetylene (indirect Fick) method of Grollman.

The blood flow through the kidney or liver in man, as well as that through the coronary circulation, has also been ascertained by this method, blood being obtained, respectively, from the renal vein, hepatic veins (p 336) or the coronary sinus.

The indirect Fick method of Douglas and Haldane based upon the CO₂ arterio-venous difference

The CO₂ tension of the arterial blood is obtained from the tension of CO₂ in the alveolar air. The CO₂ tension of the mixed venous blood is obtained by breathing air mixtures containing different percentages of CO₂ from a series of bags, and holding the breath for a sufficient length of time to allow the mixture in the lung-bag system to come into equilibrium with the venous blood. That gas mixture of which two samples taken a short interval apart have practically the same CO₂ percentage is assumed to be in equilibrium with the venous blood. The whole procedure must not take longer than the time of a single circulation (12 seconds), i.e., the blood which has left the lungs must not be allowed time to return before the second sample is obtained, otherwise the CO₂ tension of the mixed venous blood would be raised artificially as a result of the absorption of CO₂ from the air mixture, and the results be vitiated. That is, the calculated CO₂ content of the mixed venous blood would be too high and the value for the cardiac output, in consequence, too low. Blood passes through the coronary circuit in about 10 seconds so the value for the cardiac output does not include the blood of the coronary system.

Three (or four) bags of 30 liters capacity are first filled with mixtures of air and CO₂. The respective mixtures are made up with proportions of CO₂ which increase in successive bags by 0.5 per cent. Thus, the mixture in bag number one contains approximately 6.5 per cent CO₂, bag two, 7.0 per cent, bag three, 7.5 per cent. The subject makes a maximal expiration into room air and then a maximal breath is taken from bag one. These two respiratory movements are repeated three times in order to ensure thorough mixing in the lungs, and to wash out the alveoli with the gas mixture in the bag. Each expiration is made into room air and each inspiration is taken from the bag. After the final inspiration the breath is held for 2 seconds and then a sharp partial expiration (of about 1600 cc) is made down the Haldane alveolar air tube (p. 363) and a specimen taken for CO₂ analysis. Due to the dilution with the lung air (which has a lower CO₂ percentage) this sample will be lower in CO₂ than that of the original bag mixture. The subject continues to hold his breath after this short expiration, and 6 seconds later a full expiration is made in order to expel the remainder of the air from the lungs. A second sample is taken for analysis. If this should show the same CO₂ percentage as the first it would then be inferred that in the interval the venous blood had neither given up nor absorbed any CO₂ from the lungs. In other words, the venous blood and alveolar air would be in equilibrium, and the tension of the CO₂ in the sample the same as that of the venous blood. Usually, however, the second analysis gives a higher value, CO₂ must have been given off by the venous blood, and the CO₂ tension of the latter is, therefore, higher than that of the mixture in bag 1. The same procedure is then repeated with the richer mixture in bag 2 (7.0 per cent CO₂). The second of the two analyses may still be higher than the first. If so, the third bag is employed when the second sample will either be the same or slightly lower than the first sample. The venous CO₂ tension in the latter case would be between that of the second samples of the second and third observations. On the other hand, if after breathing from bag 2 the second sample has a lower CO₂ percentage than has the first, CO₂ must have passed from the lungs to the venous blood (see table 22). With the use of the third bag (7.5 per cent CO₂) the second sample will also be lower, but the difference between it

and the first sample will be greater than when the mixture in bag 2 was employed. This method of "straddling" shows that the gas mixture which would be in equilibrium with the venous blood must have a CO₂ percentage somewhere between those of the second samples in the first two observations (between 6.40 and 6.63 per cent in table above). A mixture having this percentage of CO₂ is then employed. When this is inspired and the usual procedure repeated, it is to be expected that there will be practically no difference between the first and second alveolar air analyses.

The CO₂ tension of the mixed venous blood is readily calculated from the CO₂ percentage of the gas sample in equilibrium with it (ch. 32). The volumes per cent of CO₂ in the arterial blood and in the mixed venous blood are then obtained from the CO₂ dissociation curve, the lower (oxygenated) curve being employed for this purpose (p. 396).

The total CO₂ elimination per minute is determined just before the re-breathing procedures by the Douglas bag method (ch. 45). Then, if the total CO₂ elimination is, say, 208 cc per minute and the arterio-venous CO₂ difference is 5 volumes per cent.

$$(208/5) \times 100 = 416 \text{ liters, cardiac output per minute}$$

METHODS INVOLVING THE USE OF A FOREIGN GAS

Nitrous oxide (N₂O) and *ethyl iodide* (C₂H₅I) have been employed in the past. A method involving the use of *acetylene* (C₂H₂) was introduced in 1929 by Grollman. This gives the most accurate results and has largely superseded the older methods.

If a subject breathes an inert foreign gas (i.e., one which becomes dissolved in the plasma but does not combine with the hemoglobin, lipid or other constituents of the blood) and the quantity which has been absorbed in a given time be known, as well as the coefficient of solubility of the gas in plasma, then the quantity of blood which has passed through the lungs can be calculated. This, in general, is the principle upon which the acetylene or any other foreign gas method is based.

The acetylene method (Grollman). The subject rebreathes a mixture of acetylene (20 to 25 per cent) and air from a rubber bag (3 liters capacity) until the lung-bag system contains a homogeneous mixture. A sample of the mixture is taken and the percentages of acetylene, nitrogen and oxygen determined. The subject continues to rebreathe for 5 seconds. A second sample is then taken for analysis. The entire rebreathing procedure should not exceed the time of a single circulation (12 seconds) in order that no acetylene shall be returned to the lungs, and also that the arterio-venous oxygen

TABLE 22

	CO ₂ PER CENT AIR IN BAG	CO ₂ PER CENT IN ALVEOLAR AIR		PRESSURE OF CO ₂ IN VENOUS BLOOD IN RELATION TO ALVEOLAR AIR
		First sample	Second sample	
(1)	6.50	6.38	6.40	+
(2)	7.00	6.68	6.63	—
(3)	7.50	6.87	6.73	—

difference shall not be altered from that existing prior to the rebreathing period

The volume of the lung bag system is reduced during the interval between the taking of the two samples, since the volumes of acetylene and oxygen absorbed are together greater than the carbon dioxide eliminated. Inasmuch as nitrogen takes no part in the respiratory exchanges, an increase in its percentage will be proportional to the reduction of the volume of the mixture in the lung bag system. The relative gas volumes in the system at the times of the first and second samples, which may be designated V_I and V_{II} respectively, are therefore calculated from the nitrogen percentages at these times, V_I , V_{II} , N_{II} , N_I

If $(C_2H_2)_I$ and $(C_2H_2)_{II}$ be used to designate the percentages of acetylene in the first and second samples, respectively, then the relative volumes of acetylene in the samples will be $V_I(C_2H_2)_I$ and $V_{II}(C_2H_2)_{II}$, and the volume of acetylene absorbed will be $V_I(C_2H_2)_I - V_{II}(C_2H_2)_{II}$. Knowing the average concentration of acetylene—

$$\frac{(C_2H_2)_I + (C_2H_2)_{II}}{2} = (C_2H_2)_{\text{aver}}$$

in the lung bag system in the interval between the taking of the two samples, and the coefficient of solubility of the gas, then the quantity absorbed per liter of blood is,

$$\frac{740 (C_2H_2)_{\text{aver}}}{100} \times \frac{B - 48.1}{760}$$

740 = the number of cubic centimeters of acetylene which are dissolved by 1 liter of blood at body temperature when the tension of the gas is 760 mm Hg

B = barometric pressure

48.1 = tension of water vapor in the lungs

Then

Quantity of C_2H_2 absorbed during observation
Quantity of C_2H_2 absorbed per liter of blood

$$= \frac{[V_I(C_2H_2)_I - V_{II}(C_2H_2)_{II}] \times [760 \times 100]}{740 (C_2H_2)_{\text{aver}} \times (B - 48.1)}$$

= Blood flow through lungs, i.e. cardiac output during rebreathing

The blood flow through the lungs, thus obtained, is affected by the experimental procedure itself, namely, the rebreathing. In order to obtain the cardiac output for the period immediately preceding the rebreathing period, the arterio-venous oxygen difference and the total oxygen consumption must be determined. Now, the arterio-venous oxygen difference during the actual experiment will be the same as that immediately preceding the period of rebreathing since this is shorter than the time of a single circulation. The oxygen ab-

sorption during the rebreathing period is obtained in a manner analogous to that by which the acetylene absorption is obtained. The volume of oxygen absorbed divided by the quantity of blood in liters passing through the lung during rebreathing gives the arterio-venous oxygen difference i.e., the quantity of oxygen absorbed by each liter of blood during its passage through the lungs. Thus—

A V oxygen diff

$$\begin{aligned} &= \frac{[V_I(O_2)_I - V_{II}(O_2)_{II}] \times 740(C_2H_2)_{\text{aver}} (B - 48.1)}{V_I[(C_2H_2)_I - V_{II}(C_2H_2)_{II}] \times [760 \times 100]} \\ &= \frac{(O_2)_{\text{diff}} (C_2H_2)_{\text{aver}} \times (B - 48.1) \times (0.00974)}{(C_2H_2)_{\text{diff}}} \end{aligned}$$

The factor 0.00974 is derived by combining the constants 760 and 100, and 740—

$$\left(\frac{740}{760 \times 100} = 0.00974 \right)$$

The total oxygen consumption is determined in the usual way just prior to the commencement of the rebreathing period. Then

Total O_2 consumption

Arterio-venous oxygen difference

= cardiac output in litres per minute

PHYSICAL METHODS

The *ballistocardiographic method* was originated by Henderson some years ago and has been elaborated and modified in recent years by Starr and his colleagues. The cardiac output is calculated from the record made by the recoil of the body caused by oppositely directed movements of the heart and by the impact of the blood during systole. It involves the basic principle that "every reaction has an opposite and equal reaction." The apparatus or ballistocardiograph consists of a table suspended from the ceiling by wires and braced to prevent any but a horizontal movement in the long axis of the body (fig. 26.1). The patient lies supine on the table with his feet braced against a foot-board. The movements of the table are opposed by a strong spring and magnified some 8,000 times through an optical recording system. The apparatus is calibrated by subjecting the table to a static force of 280 grams which causes the displacement by 1 cm. of the spot of light on the photographic surface. The normal ballistocardiogram shows three principal waves H, I, and J, inscribed during systole. Wave H is due to a small headward movement of the body and is caused by the movement feetward of the heart and of the blood within it, in the isometric period of systole (fig. 26.2). Wave I is due to a sharp recoil of the body feetward due to the ejection of blood into the aorta. The J

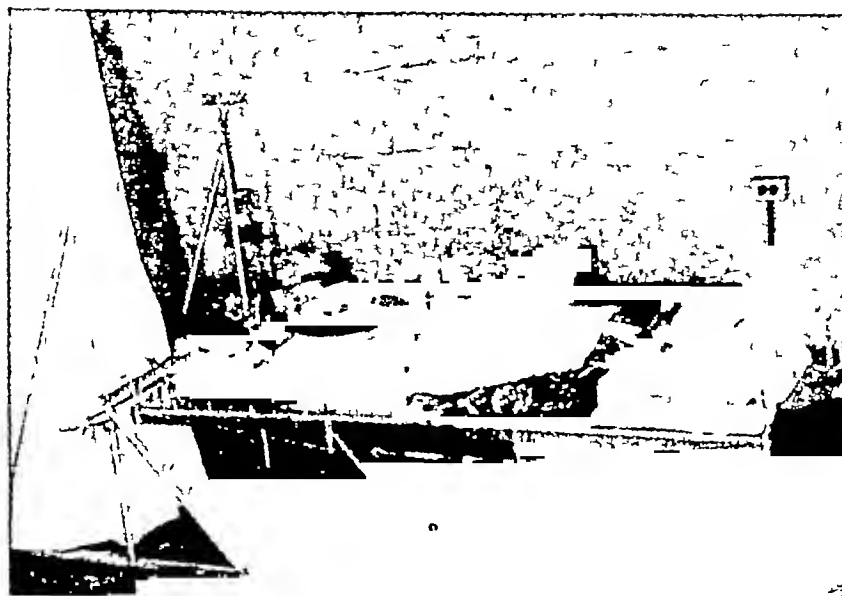


FIG 26 1 The ballistocardiograph in use (after Starr)

wave is the result of a headward movement caused by the recoil of the aorta as the blood flows down the descending aorta. The stroke volume is calculated from the areas of the waves, I and J, and by applying the following formula

$$\text{Stroke volume} = 7\sqrt{[I + J]AC^2/3}$$

where I and J are the areas of the waves so designated, \bar{A} is the diameter of the aorta (calculated from age and surface area according to Bazett's data) and C the duration of the cardiac cycle, the minute volume is obtained by multiplying the value of the stroke volume by the pulse rate.

Measurements of the cardiac output by the direct Fick method (p 267) give values 18.5 per cent higher than those obtained by the ballistocardiographic method as just described, but when the figure for aortic cross section is obtained by a roentgenological method this difference is reduced to 3.5 per cent (Cournand and his associates)

The method is very largely empirical, it being impossible to make a true estimate of the forces concerned

The blood pressure (pulse-pressure) method (as employed by Bazett and his associates) The stroke volume of the heart can be calculated if the following are known, (1) the volume of the arterial vessels at the end of the diastole, (2) the outflow from the arterial system during the cardiac cycle and (3) the distensibility of the arterial walls. The volume of the arterial system is calculated in sections—from commencement of aorta to the third part of the subclavian artery—(\bar{V}_1) the descending aorta, with its iliacs and other branches except those nearer the heart than the subclavians (\bar{V}_2), the sections between the subclavian and the brachial at the elbow and including the other vessels

of the upper part of the body with the same pulse wave velocities (\bar{V}_3) and finally, the vessels of the lower limbs together with the vessels of the arms, splanchnic area and elsewhere having similar pulse wave velocities (\bar{V}_4). The values for \bar{V}_1 and \bar{V}_2 are obtained from tables giving the measurements in cadavers of different heights, surface areas and ages. The volumes \bar{V}_3 and \bar{V}_4 are assumed to vary with size of body but not with age, and are purely empirical, being based upon the ability of the equations to give results which agree with those obtained by the acetylene method

The distensibility of the arterial walls is determined from the pulse wave velocities (p 183) in the different sections, recording tambours being placed upon the apex beat, subclavian, brachial, femoral, and dorsalis pedis arteries

The outflow during diastole is determined from the change in volume of the arterial vessels from the peak of the diastolic wave to the end diastole. The volume change is estimated, in turn, from Bramwell and Hill's modification of Moen's equation, namely, *that the*

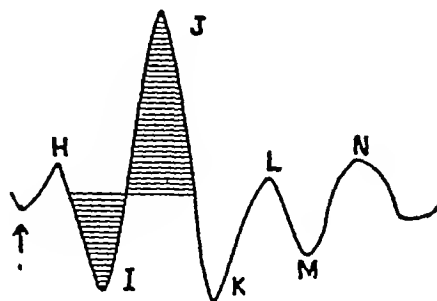


FIG 26 2 Ballistocardiograph record (semidiagrammatic) Calculations made from hatched areas (After Starr)

percentage of volume change per millimeter of mercury rise in pressure = $12.7/v^2$ where v is the pulse wave velocity in meters per second. The outflow from the arterial system during diastole is then calculated from the equation

$$V_d = \frac{12.7(Z - D)}{100} \left(\frac{\bar{V}_1}{v_1^2} + \frac{\bar{V}}{v_2^2} + \frac{\bar{V}_3}{v_3^2} + \frac{\bar{V}_4}{v_4^2} \right)$$

where V_d is the outflow of blood during diastole, Z the diastolic pressure, and D the pressure at the end of diastole, \bar{V}_1 to \bar{V}_4 the volumes of the arterial sections already described and v_1 to v_4 the pulse wave velocities in these sections.

The outflow during systole (s period) must, of course, be taken into account. It is related to and derived from the volume leaving in the diastolic period (V_d) and calculated from the ratio of the mean pressures in the two periods. Thus,

$$V_s = V_d \times s/d \left(\frac{M_s - a}{M_d - a} \right)$$

where V_s is the volume of blood leaving the arterial system during systole, M_s and M_d are the mean pressures during systole and diastole, a is a constant having usually a value of 20 when the blood pressure is expressed in millimeters of mercury. The *stroke volume* (V) = $V_s + V_d$ and the *output of the heart per minute* (V_d) = $V \times F$, where F is the pulse rate and V the stroke volume.

The results of this method have been found to agree within less than 10 per cent with those of the acetylene method.

Injection method This method, which was originally proposed by Stewart, has been developed in recent times by Hamilton and his associates. A known quan-

tity of a non-diffusible dye, e.g., T 1824, is injected rapidly into an arm vein. Samples of blood are drawn every 3 seconds and the concentration of dye determined. The dye first makes its appearance in from 20 to 25 seconds. A curve, drawn by plotting the concentrations of the dye against time, rises to a maximum and then declines in a straight slope, and a little later shows a slight secondary rise or hump. The latter indicates recirculation of the dye. Extrapolation of the curve from this point to zero will give the time required for the total quantity of blood which contains dye to be swept away had no recirculation occurred. The mean concentration of the dye over this period of time is calculated.

The cardiac output can then be arrived at from the formula, $F = 60A/TC$, where F is the cardiac output in liters per minute, A the substance injected in mg, T the duration in seconds of the concentration curve, and C the mean concentration of the dye in mg per liter. Let us say, then, that 250 mg of dye were injected, the mean concentration of the dye was found to be 10 mg per liter, and the total time required for the dye to be removed, had no recirculation occurred, is, say, 32 seconds. Then the volume of blood required to carry the total quantity of dye injected is $(\frac{250}{10} \times \frac{60}{32})$ 4.68 liters per minute (cardiac output).

Radiographic method In this method the stroke volume is calculated from the area of the heart shadows in instantaneous skiagrams, or by roentgenkymography (Keyes), taken in systole and diastole. The heart volumes in cc. in systole and diastole are calculated from the cardiac areas (after correction for distortion) in these periods by means of the following formula, Volume = 0.64 (area) 145. The difference between these values gives the stroke volume.

CHAPTER 27

THE CONTROL OF THE BLOOD VESSELS THE VASOMOTOR MECHANISMS PERIPHERAL VASCULAR DISORDERS SURGICAL SHOCK

The walls of the arterioles are composed chiefly of involuntary muscle fibers arranged in a circular fashion. Like the cardiac muscle the musculature of these vessels is supplied by two types of nerve fibers—inhibitory and excitatory. Those which cause contraction of the arteriolar musculature are called *vasoconstrictor*, those which inhibit, and in consequence cause relaxation of the muscular rings, are termed *vasodilator*. The former are therefore analogous to the cardiac accelerator (augmentor) nerves, and the latter to the vagi. Both sets together are referred to as the *vasomotor nerves*.

THE VASOCONSTRICTOR FIBERS

These were discovered in 1852 by Claude Bernard, who stimulated the cervical sympathetic nerve in the rabbit and observed constriction of the vessels of the ear. They belong to the thoracico-lumbar (sympathetic) division of the involuntary nervous system. The constrictor fibers arise from groups of nerve cells situated in the lateral horns of the spinal gray matter, extending in man from the first thoracic to the second or third lumbar segment, inclusive. All the arterioles of the body wherever situated are supplied with filaments whose ultimate source is in this relatively limited region of the central nervous system. They are distributed to the periphery in the manner elsewhere described for the thoracico-lumbar outflow in general (see also chapter 72).

The vascular nerves of the limbs, as shown by Fodd and Kramer and by Woolfard, are distributed by two distinct modes. (1) A *proximal* innervation which arises in the case of the vessels of the upper limb directly from the cervical part of the sympathetic chain—middle and inferior cervical ganglia. The fibers pass to the subclavian artery and are conveyed in a plexiform manner along the outer coat of this vessel and its branches, and into the arm along the axillary artery. The corresponding supply to the vessels of the lower limb is derived by extension from the aortic plexus in the abdomen. The fibers follow the

common and external iliac arteries into the thigh. The sympathetic fibers derived in the manner just described do not extend beyond the larger vessels of the limbs—proximal portions of the brachial and femoral. (2) A *distal* innervation which is carried to the peripheral vessels via the spinal nerve trunks (e.g., ulnar, sciatic, etc., fig. 27.1). These reach the arteries at different levels and, penetrating the vascular wall, form a nerve net surrounding the muscular coat, the highest level of this type of innervation probably overlaps the region innervated by the proximal group of fibers mentioned above. The lowest levels supply the arterioles and capillaries. It is solely through such sympathetic fibers traveling with somatic nerve trunks that constrictor impulses are conveyed to the minute vessels of the limbs. Ganglion cells are absent from the vessels of the limbs. Section of a peripheral nerve, therefore, causes complete degeneration of vasoconstrictor fibers in the area of its distribution.

The existence of the distally distributed set of vasoconstrictor fibers has an important bearing upon operations designed to denervate the vessels. *Periarterial neurectomy*, for example, in which a segment of a main artery is stripped, will interrupt fibers belonging to the proximal set but will leave the distal supply to the vessels intact. Histological examinations of the peripheral vessels of limbs which had been amputated some time after periarterial neurectomy had been performed have shown only *undegenerated* nerve filaments.

Vasoconstrictor fibers to the head and neck are conveyed from the sympathetic chain through plexuses investing the blood vessels, but also via peripheral nerve trunks (cervical and certain cranial nerves). The vessels of the abdomen and pelvis are supplied with fibers which pass along the vascular walls from plexuses surrounding the aorta and its branches.

The vasoconstrictor fibers are non-medullated, but other fine medullated fibers may be detected ramifying around the peripheral vessels. These are afferent and convey sensory impulses (pain) from the vessels as well as dilator impulses (anti-

dromic, p 276) to the vascular muscle, they travel in the mixed somatic nerves and enter the cord by the posterior spinal nerve roots. None degenerates after removal of the sympathetic chain.

Evidence has accumulated within recent years which indicates that vasoconstriction, like certain other sympathetic effects, is mediated through the liberation of noradrenaline (chap 59) at the nerve endings (see p 248).

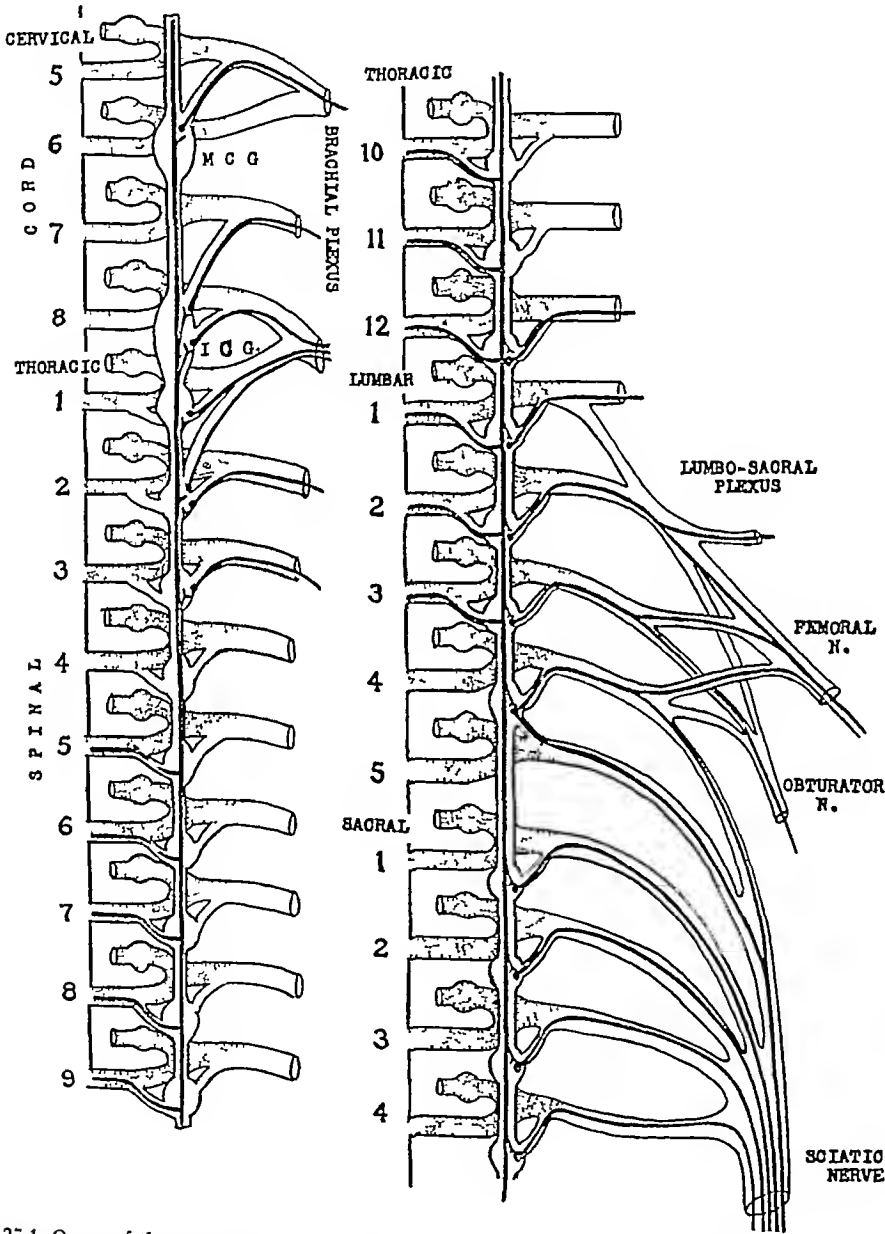


FIG 27 1 Origin of the vasoconstrictor fibers to the limbs and their distribution in the peripheral nerves M C G middle cervical ganglion, I C G inferior cervical ganglion Superior cervical ganglion and distribution of vasoconstrictors to head are not shown (see also ch 72) A R anterior spinal nerve root, P R posterior spinal nerve root Preganglionic fibers, red, postganglionic, blue

TABLE 23
Distribution of vasoconstrictors (see also chapter 72)

REGION OF BODY	ORIGINS	
	Preganglionic fibers	Postganglionic fibers
Head	1st and 2nd thoracic segments of cord (man)	Superior cervical ganglion
Upper limb and neck	2nd (and probably 1st) to 8th or 9th thoracic segments inclusive (man)	1st, 2nd and 3rd thoracic and inferior and middle cervical ganglia thence via nerves of brachial plexus
Lower limb	10th thoracic to 2nd lumbar segments	From 12th thoracic to 4th sacral ganglia thence via nerves of lumbo-sacral plexus
Thoracic aorta and branches	1st to 5th thoracic segments	Middle and inferior cervical and 1st to 5th thoracic ganglia
Pulmonary vessels	2nd, 3rd and 4th thoracic segments	
Abdominal and pelvic viscera	6th thoracic to 3rd lumbar segments inclusive	Celiac, mesenteric and hypogastric ganglia (no cell stations on vertebral chain of ganglia)

THE VASODILATOR FIBERS

The origins of the vasodilator fibers are more diverse than those of the vasoconstrictors. Dilator impulses emerge from the central nervous system by (1) the *thoracico-lumbar outflow*, the fibers of which reach a given vascular area by the same route as that travelled by the vasoconstrictor impulses, (2) the *cranial outflow* of the parasympathetic division reaching the periphery via the *chorda tympani*, *glossopharyngeal* and the *vagus* nerves, (3) the *sacral outflow or pelvic nerve*, and (4) the *posterior spinal nerve roots—antidromic impulses*.

VASODILATORS OF THE THORACICO-LUMBAR DIVISION (SYMPATHETIC)

The question of vasodilator fibers in the sympathetic has been debated, since most workers in the past doubted their existence. Several observations had, nevertheless, suggested such a distribution. Dastre and Morat, for example, observed flushing of the buccal mucosa upon stimulation of the cervical sympathetic; this observation was confirmed by others (Carlson, Langley). The phenomenon, however, was ascribed by Gaskell, Bayliss and others not to vasodilator impulses but to the liberation of vasodilator substances (CO_2 , lactic acid, etc.) from neighboring small glands of the mucosa, excited by the stimulation of secretory fibers contained in the cervical sympathetic.

On the other hand, Dale showed that after

ergotoxine, which abolishes secretory as well as vasoconstrictor and other motor sympathetic effects, stimulation of the splanchnic or of the abdominal sympathetic caused dilatation of the vessels of the intestine or of the leg, respectively. This was strong evidence for the existence of sympathetic vasodilator fibers. The action of the dilators had apparently been unmasked when the stronger vasoconstrictor effect was annulled by the drug.

Burn produced further evidence for the presence of vasodilators in the lumbar sympathetic chain of the dog. After the blood pressure had been elevated (by the infusion of adrenaline), stimulation of this nerve caused vasodilatation and a fall in blood pressure.

The sympathetic vasodilators are of two types, namely, those which bring about their effect by the liberation of acetylcholine (*cholinergic fibers*, ch 72) and those whose action is mediated by an adrenaline-like substance (*adrenergic fibers*). The distribution of each type varies with the species and in different vascular areas of the same species. In the dog and hare the muscles are plentifully supplied with sympathetic vasodilators of the cholinergic variety (Bülbring and Burn). Stimulation of sympathetic fibers to the muscle vessels of the dog after ergotoxine administration (which paralyzes vasoconstrictor fibers) causes vasodilatation, the reaction being enhanced by eserine (physostigmine) and abolished by atropine. Moreover, a vasodilator response is obtained from the stimulation of sympathetic fibers if eserine has been administered previously, the action of the vasodilators now overbalancing

TABLE 24

Summary of vasodilator fibers (see also chapter 72)

REGION	ORIGIN	COURSE
Head	Parasympathetic (cranial outflow) Thoracico-lumbar outflow	7th, 9th and 10th cranial nerves Cervical sympathetic cord
Fore limb	6th 7th, 8th cervical and 1st thoracic segments Thoracico lumbar outflow probably from same segments as those giving rise to constrictors (p 275)	Posterior roots of corresponding somatic nerves Probably same as vaso-constrictors
Hind limb	5th, 6th and 7th lumbar 1st and 2nd sacral Thoracico-lumbar outflow probably from same segments as those giving rise to constrictors	Posterior roots of corresponding somatic nerves Abdominal sympathetic chain and somatic nerves
Abdominal and pelvic viscera	Thoracico-lumbar outflow, probably from same segments as those giving rise to constrictors	Splanchnics
Penis or clitoris	Parasympathetic (sacral outflow) 2nd, 3rd and 4th sacral segments	Anterior roots and nervi erigentes (pelvic nerves)

that of the constrictors. There are only a few sympathetic vasodilators in the muscles of the cat and they are entirely adrenergic, the vasodilation caused by stimulation of the sympathetic after ergotoxine administration being unaffected by eserine or atropine. That in the cat, some other vascular area is innervated by cholinergic fibers is indicated by the experiments of Cannon and Rosenblueth, who found that the fall in blood pressure following stimulation of the abdominal sympathetic after ergotoxine was enhanced by eserine and reduced or abolished by atropine. The muscles of the monkey and of the rabbit receive no sympathetic vasodilator fibers of either kind. The skin of the dog, except over the ears, is devoid of sympathetic vasodilators. The presence of sympathetic vasodilators in the human skin is in dispute. The experiments of Lewis and Pickering seemed to have demonstrated the existence of such fibers, but Uprus, Gaylor and Carmichael in a later investigation have questioned the conclusions of the previous workers (see, however, Fatheree and Allan for affirmatory evidence).

The increase in blood flow through the muscles when the entire body is heated, or during a faint, is brought about through vasodilatation, mediated through the sympathetic (Wilkins and Eichna, Barcroft and Edholm, McMichael and Sharpey Schafer). There is a decrease of vasoconstrictor tone as well as active vasodilatation. If the vessels in all the muscles become dilated when the body is over heated, as is likely, a

great increase in vascular capacity will result, which explains the tendency to faint when the air temperature rises to an uncomfortable height.

ANTIDROMIC VASODILATOR IMPULSES

Stricker many years ago (1876-1877) reported that stimulation of the peripheral segments of the cut posterior roots of the sacral nerves caused dilatation of the vessels of the hind paw of the dog. This observation was at variance with the Bell-Magendie law which states that the posterior roots convey only centripetal impulses. Stricker's results were, consequently, not generally accepted. The question was studied afresh by Bayliss who showed in a series of carefully controlled experiments that Stricker's observation was essentially correct. Stimulation of the distal cut ends of the posterior roots of the 4th lumbar or the 1st sacral nerve, for example, was followed in every instance by dilatation of the vessels of the hind limb. Electrical, thermal and, especially, mechanical stimuli were effective. The effectiveness of the latter type of stimulus, which can be strictly localized, showed that the dilator response was actually due to the stimulation of posterior root fibers and not to the inadvertent stimulation of the anterior roots, as might occur with the use of an electrical type of

stimulus (as a result of the escape of current) Furthermore, dilator effects in the vessels of the hind limb could be produced by stimulating the central end of an afferent nerve (such as the cardiac depressor, ch 27) after the abdominal sympathetic chain had been excised This experiment shows that fibers of the posterior roots constitute the efferent limb of a depressor reflex

There still remained the possibility, nevertheless, that *efferent fibers*, leaving the cord in the posterior roots, and not the ordinary sensory fibers were responsible for the dilator effect Such a possibility was excluded, however, by sectioning the root between the ganglion and the cord and allowing time for degeneration to take place Any efferent fibers would be cut off from their nerve cells by this operation and must, therefore, undergo degeneration Afferent fibers, on the contrary, being still in connection with their cell bodies in the root ganglion would survive Following the period allowed for degeneration, stimulation of the peripheral segment of the sectioned root caused the usual vasodilator response On the other hand, when the section was made *distal* to the ganglion, stimulation of the distal stump after degeneration of the sensory fibers had occurred failed to cause vasodilatation Trophic centers for the vasodilator fibers must, therefore, lie in the posterior root ganglia The application of nicotine to the ganglion, which blocks transmission at the synaptic junctions, did not abolish the vasodilator response when the posterior root was stimulated on the central side of the ganglion This fact indicates that the dilator fiber does not synapse within the ganglion All the evidence points to the vasodilator fibers being derived from the ordinary unipolar ganglion cells, i e, that they are identical with the afferent fibers of the peripheral nerves and posterior roots In other words, the vasodilator impulses are conveyed along the fiber in a direction opposed to that in which the ordinary sensory impulses travel For this reason they have been termed *antidromic* (running against) by Bayliss and Langley Sherrington, from histological evidence, had also concluded that the posterior roots were free from efferent fibers He sectioned the roots in monkeys and cats proximal to the ganglia and, after time had been allowed for degeneration to occur, examined the peripheral stumps for degenerated fibers, none was found

Foerster, more recently, obtained cutaneous vasodilatation in man by stimulation of the poste-

rior nerve roots at various levels of the cord The vasodilatation had a segmental distribution

According to Bayliss and Head, the sensory fibers which transmit the vasodilator impulses are those which subserve the sensation of pain, and it was shown by Bayliss that in the case of the limb the antidromic impulses pass mainly to the vessels of the skin, few, if any, to vessels of the muscles It appears that the vasodilator fibers of the posterior roots are distributed mainly if not entirely to the cutaneous and visceral vessels (the fibers to the latter travelling in the splanchnics) Bayliss observed that the limb deprived of skin showed only a slight increase in volume after stimulation of the appropriate posterior roots The dilators to the vessels of skeletal and cardiac muscle, on the other hand, are derived from the sympathetic (pp 275 and 326) The vessels of skeletal muscle are therefore dependent upon the sympathetic for the transmission of both dilator and constrictor impulses As compared with the skin and viscera, however, the muscular vessels rely to a greater extent upon metabolic products, carbon dioxide and lactic acid, for the control (p 291) of their calibers The vasodilatation which occurs in glands accompanying their secretion is also probably due to a large extent to the direct action of metabolites upon the vascular walls

Antidromic vasodilator fibers homologous with those of the posterior roots are believed to exist also in certain cranial nerves, e g, the trigeminal, which carried dilator impulses to the face and tongue (lingual) The conception of an antidromic vasodilator mechanism has far-reaching implications and has helped very greatly toward an understanding of several clinical phenomena hitherto inexplicable Of these may be mentioned the vascular changes in certain diseases affecting the sensory side of the nervous system and the cutaneous lesions arising in tabes, especially along the course of the lightning pains The blisters which occur in herpes zoster have been shown by Head and Campbell to be associated with lesions of the cells of the ganglia of the posterior nerve roots

Within recent years the question of antidromic transmission of vasodilator impulses has been questioned by some observers who claim to have demonstrated the presence of efferent fibers in the posterior roots, and that they are responsible for the vasodilator effects The Japanese school, headed by Ken Kure, has referred to these fibers as constituting a "spinal parasympathetic system" This conception has not gained

general acceptance. More recently, however, Kahr and Sheehan obtained evidence for the existence of efferent fibers in the posterior roots of cats. When the posterior roots from 12 thoracic to 2 lumbar were sectioned between the ganglion and the cord and time allowed for degeneration to occur, undegenerated fibers were found in the proximal stumps and degenerated fibers in the distal stumps. These, it was concluded, must have been derived from nerve cells situated within the central nervous system, and studies of sections of the cord revealed changes in the Nissl bodies of cells situated in the lateral and anterior horns ("retrograde" degeneration, p. 913). But, though the existence of efferent fibers in the posterior roots should be demonstrated conclusively by histological methods, it does not necessarily follow that they convey vasodilator impulses. The physiological experiments of Bayliss cited above furnish strong evidence that vasodilator impulses are transmitted antidromically in the posterior roots. Even if efferent fibers made connection with certain cells in the root ganglia (as is maintained by Kure, but which remains unproved) Bayliss' experiments could be explained, only in part, otherwise than by the antidromic hypothesis. Nevertheless, the conception of antidromic impulses has been a difficult one for physiologists to accept because it contradicts the Bell-Magendie law (ch. 66). It is not possible to discuss the question here except in the most cursory manner. The reader is referred to a paper by Bishop and associates and to a recent article by Barron and Matthews. The former observers consider that the vasodilator fibers of the posterior roots, though having their cell stations in the posterior roots are efferent in function, and analogous to similar fibers in the vagus. Barron and Matthews find that certain fibers of the posterior roots, which from degeneration experiments apparently have their trophic centers within the cord, are in reality derived from posterior root ganglia situated at a lower level. That is, the sensory fiber after entering the cord and ascending for a short distance, gives off collaterals which emerge in the posterior root of a higher spinal segment. When the posterior root is sectioned between the posterior root ganglion and the cord, these collateral fibers must, of course, undergo degeneration distal to the point of section. The presence of such fibers may explain the findings of Kahr and Sheehan mentioned above. It is clear that the results of further investigation alone can lead to a complete elucidation of the nature of the vasodilator fibers in the posterior roots.

THE PELVIC NERVE (NERVUS ERIGENS) AND THE PHENOMENON OF ERECTION

The pelvic nerve is composed of fibers which leave the cord in the anterior roots of the 2nd, 3rd sacral nerves, and sometimes the 1st and 4th, it conveys dilator fibers to the vessels of the penis

or clitoris. The erectile tissue of these organs is composed of cavernous blood sinuses whose walls contain involuntary muscle. To reach these spaces the blood passes through arterioles and capillaries. The outlets from the sinuses are guarded by rings of involuntary muscle. Dilator impulses cause arteriolar and capillary dilatation coincident with inhibition of the involuntary muscle in the walls of the sinuses, and excitation of the muscle guarding their outlets. These effects cause dilatation of the vascular spaces, a high pressure within them, and, as a consequence, hardening and erection of the organ. There is also evidence that arteriovenous anastomoses (p. 319) also exist which open up to increase the blood flow into the erectile tissue. The narrowing of the venous outlets impedes the outflow from the sinuses only until the blood pressure within their cavities is raised to a certain height, then the outflow from the organ equals the inflow and the velocity of flow through the erectile tissue becomes greatly increased. This is evidenced by the rise in temperature and the bright arterial color of the blood flowing along the dorsal vein of the penis. The arterioles, capillaries and walls of the sinuses are also furnished with constrictor nerves, derived from the prostatic plexus (ch. 72), when stimulated these nerves, by reducing the blood flow through it, cause shrinkage of the organ. The afferent pathway for the reflex of erection is through the pudendal nerves.

The pelvic nerve also sends dilator fibers to the vessels of the rectum, descending colon and bladder.

THE VASOMOTOR CENTERS

The constrictor and dilator vascular effects are controlled by centers—the *vasoconstrictor* and *vasodilator centers*—situated in the floor of the 4th ventricle of the medulla. The constrictor center is also connected to a subsidiary center (or group of centers) in the cord. This is constituted of those cells in the thoracicolumbar region of the cord already mentioned (p. 273). Ranson and Billingsley explored the floor of the 4th ventricle by means of a needle electrode and located two points which when stimulated caused respectively a rise or a fall in blood pressure of from 30 to 40 mm Hg. The former point, which probably represents the vasoconstrictor center, occupies the apex of the ala cinerea or the fovea inferior. The second point—the vasodilator center—lies just lateral to the obex. Both centers are bilaterally repre-

sented. However, these areas do not represent the highest centers of the vasomotor system. More recent work indicates that the latter are situated in the hypothalamus, and even in the cerebral cortex (premotor and orbito-frontal areas, chaps 68 and 72).

VASOMOTOR TONE AND ITS REGULATION

The constrictor center exhibits tone. It is generally stated that dilator tone is absent, but Bayliss has shown that this exists to a slight degree, under certain circumstances at any rate. Dilator tone is much more difficult to demonstrate since constrictor and dilator fibers in most instances run together and the effects of the former mask those of the latter. The tone of the vasoconstrictor center may be demonstrated by sectioning the cord in the lower cervical region. This interrupts the stream of vasoconstrictor impulses passing from the medullary to the spinal centers, the vessels dilate and the blood pressure falls. After a time, however, the blood pressure rises again, the spinal centers exhibit their inherent power of autonomous action, and assuming the duties hitherto exercised by the medullary centers, restore the vessels to their previous state of tonic constriction. The time required for the vessels to regain their tone after section of the cord varies considerably in different species (see spinal shock, ch 65).

The high degree of vasoconstrictor tone which is normally maintained is shown by the fact that section of the splanchnics doubles the flow in the vessels of the denervated area (Burton-Opitz). A corresponding increase in the flow through the femoral artery after removal of the lumbar sympathetic has been demonstrated by Herrick, Essex and Baldes, the greater flow persists for several months.

After the tone resulting from cord section has been restored it falls again if the splanchnics are sectioned, but after a time a certain degree of tone is regained. This resides in the vascular muscle itself—*peripheral tone*. Apparently, a long period is required for the development of the intrinsic arteriolar tone. Essex and his associates found that, nearly eleven months after its denervation, the flow in the dog's femoral artery was double that in the opposite femoral. But nine years after the operation, the flow was almost equal on the two sides. Examination of the small vessels of the denervated side showed pronounced hypertrophy of the muscular coat. The vessels of this

side were also especially susceptible to the constrictor action of adrenaline (see ch 59).

The tone of the vasomotor center is dependent (1) upon afferent nerve impulses received from various organs and regions of the body as well as from other nervous centers (cerebral cortex, respiratory center etc., see p 287) and (2) upon the chemical composition of the blood.

Vasodilator tone also has been clearly demonstrated (see p 288).

VASOMOTOR REFLEXES

Vasomotor reflexes can be elicited by the stimulation of practically any afferent nerve—somatic or visceral.

VASCULAR REFLEXES RESULTING FROM THE STIMULATION OF SOMATIC NERVES

Stimulation of the central end of a nerve such as the *sciatic*, the *median* or a sensory *cranial* nerve may result in either a rise or a fall in the arterial blood pressure according to the strength and type of the stimulus employed. An elevation or depression of the blood pressure brought about in this way is spoken of, respectively, as a *pressor* or a *depressor reflex*. The components of the reflex arc upon which the responses depend are, (1) afferent fibers in the peripheral nerve, (2) the vasomotor centers, and (3) the efferent vascular nerves, i.e., the vasoconstrictors or vasodilators. In order to elicit the pressor reflex, a stimulus much stronger than that necessary to provoke the depressor response must, as a rule, be applied, that is, one which would elicit pain in a conscious animal. Stimulation of the cornea which is supplied liberally with pain fibers gives a definite pressor response. In a series of experiments by Ranson and Billingsley, the pressor response ranged from 8 to 45 mm Hg and the depressor from 4 to 22 mm Hg (fig 27.2). In the elicitation of either reflex the magnitude of the response is apparently dependent upon the number of afferent fibers involved. For example, stimulation of various nerves of the brachial or lumbar plexus caused practically equivalent depressions or elevations in the blood pressure when the number of afferent fibers in the respective nerves was taken into account. It has also been shown by Martin and Stiles that similar reflex effects evoked simultaneously from separate afferent nerves are summed, dissimilar reflexes are mutually antagonistic.

It has been thought that these contrary effects upon the blood pressure were dependent upon two corresponding types of specific afferent fibers—pressor and depressor—in the peripheral nerve. The fact that stimuli of different intensities and frequencies caused opposite effects seemed in itself to imply the presence of two sets of fibers, one with a higher threshold than the other. Weak, slowly repeated stimuli are more likely to cause a fall, strong, rapidly repeated stimuli a rise in blood pressure. Other observations seemed to point to the existence of two separate sets of fibers. After cooling the nerve, or sectioning it and allowing a

certain period to elapse, the depressor effect could be obtained but not the pressor. The pressor fibers were presumed to degenerate before the depressor. Again, depressor reflexes are obtained more readily from cranial nerves (e.g., aortic nerve), whereas the usual response to stimulation of the central end of a spinal sensory nerve is a pressor response, which suggested that different nerves contain the respective types of afferent fibers in varying proportions. The work of Ranson and Billingsley indicates that the pressor fibers are not specific but are identical with those which subserve protopathic sensibility (pain, extremes of temperatures, ch. 63), whereas the depressor fibers correspond to those which transmit sensations of touch or warmth. According to Ranson the fibers transmitting *pressor* effects are unmyelinated and enter the cord in the lateral division of the posterior roots. Their intraspinal course is by short internuncial fibers in the tract of Lissauer, the nerve cells lying in the gray matter of the tip of the posterior horn. An experimental lesion involving this pathway on both sides of the cord in the lumbar region, or of the lateral divisions of the posterior roots, was shown by Ranson and his associates to abolish the pressor reflex usually obtained by stimulation of the sciatic. The depressor reflex remained unaffected. The intraspinal part of the *efferent* limb of the pressor reflex lies in the anterior or the lateral column of the cord. The impulses (vasoconstrictor), as mentioned elsewhere, leave the cord by the white rami. In support of the conclusion that the pressor fibers are identical with those mediating protopathic sensibility, Bayliss and Head observed that after section of a peripheral nerve, the pressor reflex (but not the depressor) could be elicited from the regenerating nerve at the time when protopathic sensibility was returning. The association of pressor responses with painful sensations points to their being an integral part of the defensive mechanisms (nociceptive reflexes, adrenaline liberation, etc.).

After entering the cord by the posterior roots the *depressor* impulses ascend in that part of the lateral column occupied by the spinothalamic tracts. But they are not identical with the fibers of the latter since they end in the medulla (vasodilator center). Bilateral section of this part of the lateral column abolishes the depressor reflex; the pressor reflex becomes more pronounced, which suggests that ordinarily it contains a masked depressor element. The intraspinal *efferent* limb of the depressor reflex (vasodilator) is unknown.

It must be mentioned that not all the pressor or depressor impulses ascend to the medulla but that reflex arcs exist which have their centers within the cord, both vasoconstrictor and vasodilator reflexes can be elicited in an animal whose cord has been divided in the lower cervical region a short time previously.

Pressor and depressor reflexes can be readily



FIG. 27.2 Effect of stimulating somatic nerves upon arterial blood pressure. Upper tracing shows rise in carotid pressure during faradic stimulation of the central end of the brachial nerve of cat. Lower tracing shows fall in blood pressure as a result of stimulating the central end of the sciatic nerve. (After Ranson and Billingsley)

elicited in man by peripheral nerve stimulation. Warmth, for instance, applied to the feet or other part of the body causes vasodilatation in other parts remote from the point of application and a fall in pressure may result. If, however, the applied temperature is raised to the point where it becomes painful, reflex vasoconstriction occurs. The application of cold also usually produces the latter effect. The vasodilatation in one extremity, resulting from the immersion of another part in warm water, is not due to the stimulation of afferent endings in the heated tissues, but to the warmed blood acting upon a nervous center, through this and the efferent nerves the vasodilatation is brought about. Gibbon and Landis, for example, showed that the vascular response was abolished by occluding the circulation of the heated member and that it could be produced in the hands of a subject with complete transection of the cord by warming the feet. The response to cold is due in part to afferent impulses and in part to the effect of the altered temperature of the blood returned from the cooled member upon the nervous center. A change in temperature (a rise or fall) of the tissues of from 0.01 to 0.04°C is sufficient to induce the vascular response. Other interesting reflex vascular reactions must be mentioned. Mudd and Grant showed that a draught directed to the bare back or arm caused constriction of the vessels of the pharynx and nasal mucosa. Constriction of the vessels of the bronchial mucous membrane as a result of the application of cold to a remote part of the body has also been observed by means of the bronchoscope. On the other hand, deep inspiration (active or passive) causes reflex constriction of the cutaneous vessels, a slight fall of blood pressure and slowing of peripheral blood flow, as measured in the finger. The cutaneous vasoconstriction appears to be related mainly to the volume of air inspired, the receptors for the reflex response are probably situated in the lungs. The fall in blood pressure is attributed to reduced cardiac output and is not the cause of the reflex vasoconstriction.

From the results of experiments upon animals it is to be expected that in the human subject a painful stimulus applied to a somatic nerve will be followed by a pressor response. The excitation of psychic centers and also the liberation of adrenaline are additional factors which play an important part in the pressor response resulting from painful stimuli. On the other hand, stimulation of the mesentery, peritoneum and abdominal viscera or

of certain regions such as the anus, vagina and spermatic cord is usually followed by a fall in blood pressure. Bayliss demonstrated a fact of some practical importance, namely, that pressor reflexes become depressor in character under chloroform anesthesia, this reversal was not observed when ether was employed. Strychnine tended to annul the effect of chloroform and to convert depressor reflexes into those of the pressor type.

VASCULAR REFLEXES MEDIATED THROUGH AFFERENT FIBERS OF THE VAGUS

In the section dealing with the reflex control of the heart rate it was mentioned that the vagus contains afferent fibers which terminate in the aortic arch and heart. Stimulation of these fibers also produces alterations in the caliber of the blood vessels and, as a result, changes in blood pressure. In some instances, electrical stimulation of the cerebral stump of the severed vagus is followed by a rise in blood pressure (pressor reflex), more usually, however, a fall in pressure occurs (depressor reflex).

Vagopressor reflexes

The pressor fibers of the vagus are stimulated by a fall in the pressure of blood in the great veins emptying into the right auricle. McDowall, confirming an old observation of Pavlov's, found that when the venous pressure was lowered by hemorrhage (or by the intravenous injection of alcohol or of histamine) no change, or a moderate decline in the arterial pressure, occurred so long as the vagus nerves were intact. When, however, these nerves were severed, a fall in arterial pressure occurred if this had been unchanged prior to the nerve section, or a further fall resulted if the arterial pressure had been already lowered simultaneously with the fall in venous pressure. These observations have been confirmed by Anrep and Segall. Cocamization of the auricle has the same effect as vagotomy. It is concluded from these results that the fall in venous pressure exerts an influence upon afferent vagal endings situated in the right auricle, messages ascend to the medullary centers and cause a generalized vasoconstriction. This reflex is probably responsible in part for the vasoconstriction which occurs as the usual response to hemorrhage and surgical shock (p. 301), both of which conditions are associated with a fall in venous pressure.

It has also been shown by McDowall that a rise in venous pressure considerably above the normal

level calls forth a vagopressor reflex, as evidenced by the vasoconstriction which occurs in a limb connected to the body solely by its nerves, when saline is injected into the inferior vena cava. This reflex coincides with and is supplementary to the Bainbridge reflex (cardiac acceleration, p 246). It antagonizes the depressor reflex elicited from the aortic nerve or carotid sinus, and McDowall suggests that through its predominance over the latter reflexes (see below), the elevation of the arterial pressure is permitted to persist throughout muscular exercise.

The aortic or cardiac depressor nerve

In certain animals (e.g., the rabbit) the vagal fibers mediating the depressor response of this nerve are collected into a separate nerve which arises from the trunk of the vagus high in the neck. This branch of the vagus which is known as the *aortic or cardiac depressor nerve* was first described by Cyon and Ludwig (1866). It is purely afferent and depressor in function, when sectioned, and its central (cerebral) end stimulated, a pronounced fall in pressure occurs (fig 27.3), excitation of the cardiac end, on the other hand, causes no effect. Two factors are involved in the depressor response following stimulation of the depressor fibers (whether these are contained within the vagus itself or are segregated in the aortic nerve), (1) *slowing of the heart rate and increased force of the ventricular contraction*. The efferent fibers of the vagus of the same and of the opposite side constitute the efferent limb of the reflex arc through which this response is chiefly brought about (p 241), for its full elicitation at least one vagus must, therefore, remain intact. Reduced

tone of the cardiac accelerator nerves occurs reciprocally with the increase in vagal tone, but is much less pronounced, being both delayed in its appearance and weaker (Wang and Bouson). (2) *Vasodilatation*. The vasomotor pathways constitute the efferent limb of this reflex. Vasoconstrictor tone is reduced and vasodilator tone increased. The reflex effect upon the vessels cannot, therefore, be elicited after section of the spinal cord in the lower cervical region. The efferent fibers mediating the vascular component of the sinus reflex cross in the cervical part of the spinal cord.

The receptors of the reflex (i.e., the terminals of the aortic nerve) are situated in the aortic arch and upper part of the thoracic aorta, in the ventricles and probably also, according to Daly and Verney, in the coronary and pulmonary vessels (see fig 27.5).

The fall in pressure is due mainly to dilatation of the splanchnic vessels. The dilatation is not, however, confined to these vessels but includes those of the skin and muscles. Cardiac slowing plays a minor rôle in the production of the fall in blood pressure, for almost as great an effect can be obtained after both vagi have been cut. The depressor reflex can be elicited by mechanical or electrical stimulation of the aortic wall itself wherein the special proprioceptors are located, stretching is an especially effective type of stimulus. It was also shown by Einthoven and subsequently by several other observers that action currents ascend the nerve synchronously with the heart beats. The normal stimulus is, therefore, quite evidently the pulsatile expansion of the aortic wall, a rise in general blood pressure increasing the intensity of the stimulus, a fall in pressure causing the reverse effect.

A small structure (*glomus aorticum*) analogous to the carotid body (see footnote, p 283) is connected with a branch of a fine artery arising from the aorta beyond its arch. The chemoreceptors of this structure respond to oxygen lack (see ch 33). Reflex vasoconstriction and a rise in blood pressure follow their stimulation by anoxemia or by such drugs as cyanide, lobeline, nicotine, sodium sulphide and acetylcholine.



FIG 27.3 Fall in arterial blood pressure resulting from stimulation of the central end of the cardiac depressor (aortic) nerve. The drum was stopped in the middle of the curve and the excitation maintained for seventeen minutes. The line of zero pressure should be 30 mm lower than here shown. (From Bayliss)

THE CAROTID SINUS MECHANISM (SEE ALSO P 245)

The carotid sinus is the term applied to the slight enlargement of the common carotid artery

where it bifurcates into the internal and external carotids (fig 27 4) The dilatation usually involves, as well, the commencement of the internal carotid, or may be confined to this region The carotid sinus was shown by Hering in 1923 to play an important rôle in the regulation of the cardiac rate and arterial blood pressure¹ Compression of the carotid at its bifurcation (so as to raise the pressure within the sinus) caused a marked slowing of the heart rate, vasodilatation and a fall in blood pressure, these effects result even though mechanical stimulation of the vagus is carefully avoided Electrical stimulation of the sinus wall produced similar effects Pressure upon the common carotid some distance below the sinus (so as to reduce the blood pressure within the sinus itself) causes cardiac acceleration, vasoconstriction and a rise in arterial pressure together with, as shown by Heymans, the liberation of adrenaline The carotid sinus therefore constitutes a mechanism whereby both pressor and depressor effects are mediated The effects are brought about through the following neural mechanism

THE SINUS REFLEX ARC The afferent fibers of the reflex arc are contained in the *sinus nerve*, a branch of the glossopharyngeal This delicate filament descends between the internal and external carotids to the sinus where its fibers terminate in sensory organs (proprioceptors) situated between the connective tissue fibers in the adventitia of the sinus wall Fibers also ramify in the carotid body (*glomus caroticum*),² a small structure situated upon a branch of the occipital artery or upon a small vessel arising directly from the external carotid just above the bifurca-

¹ Sollmann and Brown were the first to describe a depressor reflex of this nature They sectioned the carotid and obtained a fall in blood pressure when traction was made upon its cephalic end They showed that the reflex was not affected by section of the vagus but was dependent upon nerve terminals in the wall of the internal carotid

² This structure known also as the *carotid gland*, *carotid body*, or *intercarotid body* is composed of rounded clumps of polyhedral cells, and possesses a rich network of capillaries of a sinusoid character Some of the cells stain brown with chromic acid (chromaffin cells, p 827) The intercarotid body, however, does not contain adrenaline, and it is not believed to be an endocrine gland It is a sensory organ containing chemoreceptors sensitive to changes in the oxygen tension of the blood and through which reflex changes in respiration are effected (see ch 33) According to Comroe, this body plays little or no part in the control of the circulation Others, however, (Heymans, Bernthal) claim that through these receptors, as through those in the aortic body, anoxemia or CO₂ excess causes reflex vasoconstriction and a rise in blood pressure



FIG 27 4 Showing the carotid sinus region in man (after Heymans) 1, common carotid, 2, carotid sinus, 3, internal carotid, 4, external carotid, 5, nerve to carotid sinus, 6, glossopharyngeal nerve

tion of the common carotid Centrally the fibers of the sinus nerve make connections with the cardio-inhibitory and vasomotor centers The efferent limb of the cardiac part of the reflex is, of course, the vagus The efferent limbs of the vasodilator and vasoconstrictor reflexes are apparently sympathetic fibers, for these reflexes are abolished by complete removal of the sympathetic chains As will be seen from figure 27 5, a nerve twig connects the sinus with the ganglion nodosum of the vagus, it also receives filaments from the pharyngeal plexus and the superior cervical ganglion of the sympathetic (see also fig 27 6)

The action currents passing along the sinus nerve have been studied by Bronk and Stella and by Partridge Electrodes were placed upon the nerve and after amplification the potential changes were recorded by means of the oscillograph At ordinary arterial pressures impulses are discharged throughout the cardiac cycle, their frequency increasing during systole and decreasing during diastole Following the chief burst of rapid impulses which synchronizes with the main wave of the pulse tracing, a second rise in frequency occurs coincident with the dicrotic wave A rise in general blood pressure increases the rate of impulse discharge as well as the number of sense organs excited The latter show slow adaptation (p 944) so that though the stimulus (distension of the arterial wall) persists the impulse discharge shows little reduction in frequency and when the pressure is very high

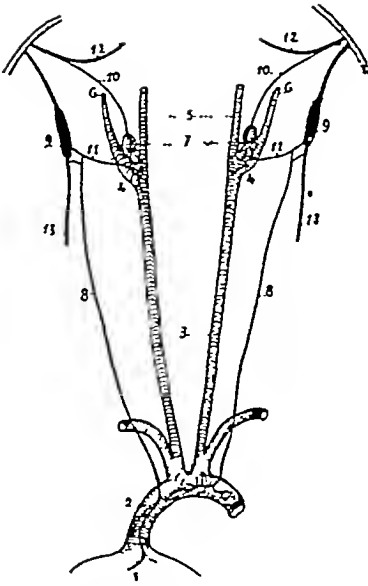


FIG 27.5 Innervation of the carotid sinus and arch of aorta. 1, heart, 2, arch of aorta, 3, common carotid, 4, carotid sinus, 5, external carotid, 6, internal carotid, 7, carotid bodies, 8, cardiac depressor nerve, 9, ganglion of vagus, 10, sinus nerve, branch of the glossopharyngeal nerve, 11, nerve branch connecting the carotid sinus with the vagus ganglion, 12, glossopharyngeal nerve, 13, vagus nerve (After Heymans)

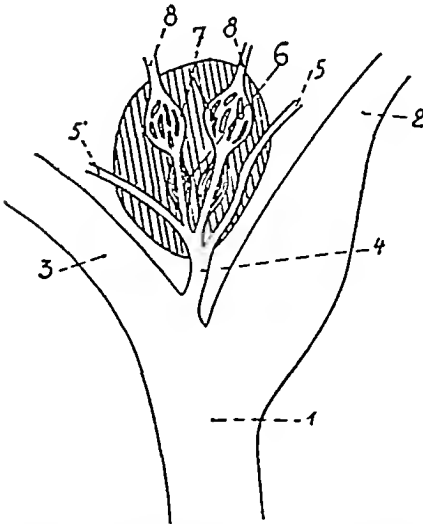


FIG 27.6 Showing the carotid body and neighboring vessels (after Heymans) 1, common carotid artery, 2, internal carotid artery, 3, external carotid artery, 4, occipital artery, 5, branches of occipital artery, 7, glomus tissue with afferent (6) and efferent vessels (8)

they extend with little reduction in rate throughout diastole. Bronk and Stella were successful in recording the impulses from a single nerve fiber (fig 27 7) The frequency of the impulses rises with increased pressure. At low pressures, impulses are discharged only during systole, but at high pressures they continue throughout diastole. The rate of impulse discharge was found to range, according to the height of the arterial pressure, from 5 to 140 per second

The sinus reflexes have been studied exhaustively by Heymans and his associates They carried out cross-circulation experiments which speak conclusively for the physiological importance of these reflexes in cardiovascular regulation (see fig 27 8) The sinus of one dog (B) was isolated from the general circulation and perfused with the blood of another animal (A) in the manner shown in the figure The nerve supply to the sinus was left intact When the arterial pressure of dog A was raised, that of dog B, recorded in the femoral artery, fell Conversely, a reduction in blood pressure of dog A caused a rise in the blood pressure of dog B In the latter instance, adrenaline liberation also occurred which was a contributory factor in the blood pressure elevation as indicated by its pressor effect upon the

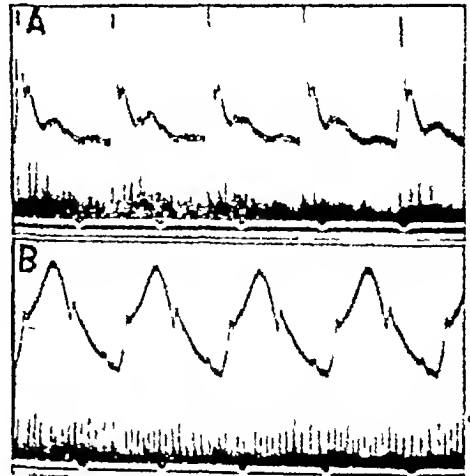


FIG 27.7 The upper curve in each record represents the arterial blood pressure registered by a membrane manometer, the lower curve of each record shows the electrical discharge from a single fiber of the carotid sinus nerve of the rabbit. In the upper record (A) the mean arterial pressure was 55 mm Hg, even at this low level a discharge of 4 impulses accompanied each ventricular systole In the lower record (B) the mean arterial pressure was 135 mm Hg, in this instance there was a more rapid and more continuous discharge from the end organ (Bronk and Stella)

circulation of dog C. These effects could not be obtained after denervation of the sinus.

The sinus and aortic nerves, or so called "buffer" nerves, constitute a mechanism of the utmost importance in controlling the arterial blood pressure and in maintaining the circulation to the brain. The rise in diastolic pressure and the increase in heart rate which occur when the body changes from the recumbent to the sitting position or from the sitting to the standing position, are apparently brought about through these nerves; they therefore play an essential part in compensating for the effect of gravity upon the circulation.

An underfilled state of the vessels, as may result from hemorrhage or shock, or any other condition which tends to cause a fall in blood pressure, will call these mechanisms into play. A generalized vasoconstriction results to adjust the vascular capacity to the reduced blood volume and thus maintain the blood pressure. Excessive elevation of the blood pressure, on the other hand, is countered by a depressor reflex (see diagram fig. 27.9). The great importance of these reflex mechanisms in hemorrhage is shown by the fact that in an animal in which all four buffer nerves have been sectioned the rapid loss of only about $\frac{1}{10}$ of the blood volume proves fatal, whereas usually a reduction in blood volume of from 35 to 45 per cent is required to cause death. Mavorson found that tilting anesthetized dogs from the horizontal to the upright position caused a sharp drop in blood pressure followed within 10 seconds by a compensatory rise. After section of both sets of buffer nerves the compensatory rise did not as a rule occur.

The cardiovascular effects are brought about through alterations in the frequency of the impulses which are constantly ascending along the aortic and sinus nerves to the cardiac and vasomotor centers. Any increase in the tension exerted upon the proprioceptors in the sinus or aortic wall causes a rise in the frequency of the afferent impulses and, as a consequence, a decrease in tone of the centers and slowing of the heart together with vasodilatation. It has been demonstrated by oscillographic methods that during a depressor reflex *efferent* impulses discharging over the cardiac accelerator and vasoconstrictor nerves are reduced in frequency or may cease. Reduction in the tension upon the vascular wall, on the other hand, by lessening the intensity of the stimulus to

the nerve endings, lowers the frequency of impulse initiation and lessens the tonic depressor effect, a rise in arterial blood pressure results. The rise in blood pressure and increase in heart rate which occur during ether anesthesia are attributed by Heymans and his associates to depression of the activity of the sinus receptors.

Several European workers (Hering, Heymans and Bouckaert, and others) have reported the occurrence of permanent hypertension in animals following bilateral section of the sinus aortic nerves. Pressures as high as 200 mm Hg lasting over a period of three years have been reported. Other investigators who have carried out similar experiments find that the hypertension so produced is not permanent in the majority of animals, but tends to return to normal after a variable period. Such a result may be due to the regeneration of the sectioned nerves or to the reflex control of the circulation being assumed by some other mechanism.

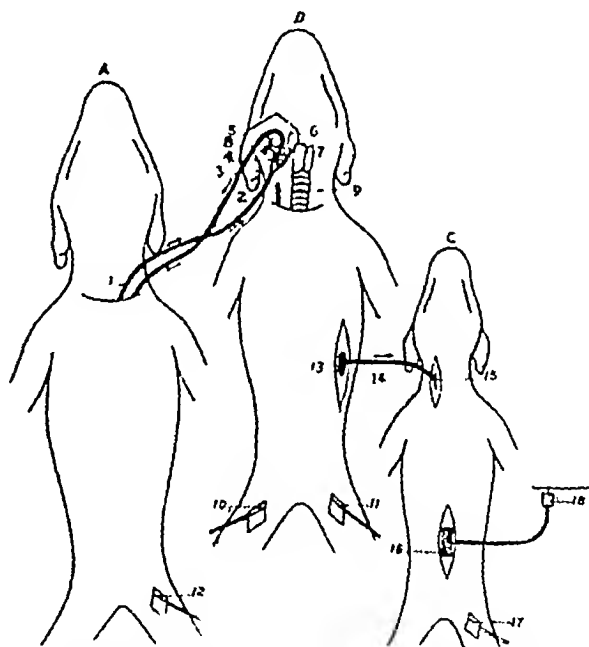


FIG. 27.8. Scheme of perfusion of the isolated carotid sinus of dog B, by dog A, and an anastomosis between the suprarenal vein of B and the jugular vein of dog C. 1, left carotid artery of dog A, 2, right carotid artery of B anastomosed with carotid of A, 3, left external jugular vein of A, 4, isolated right carotid sinus of B, 5, lingual artery of B, anastomosed with jugular vein of dog A, 6, nerve supply to carotid sinus of B. The blood from dog A flows through the carotid sinus of dog B and back to A via the lingual artery of B and the external jugular of A. 7, internal carotid, 8, facial and maxillary arteries, 9, common carotid, 10, 11, 12 and 17, femoral arteries to manometers, 13, adrenal gland, 14 and 15, suprarenal jugular anastomosis, 16, decapsulated spleen in plethysmograph, 18, piston recorder for plethysmograph (After Heymans).

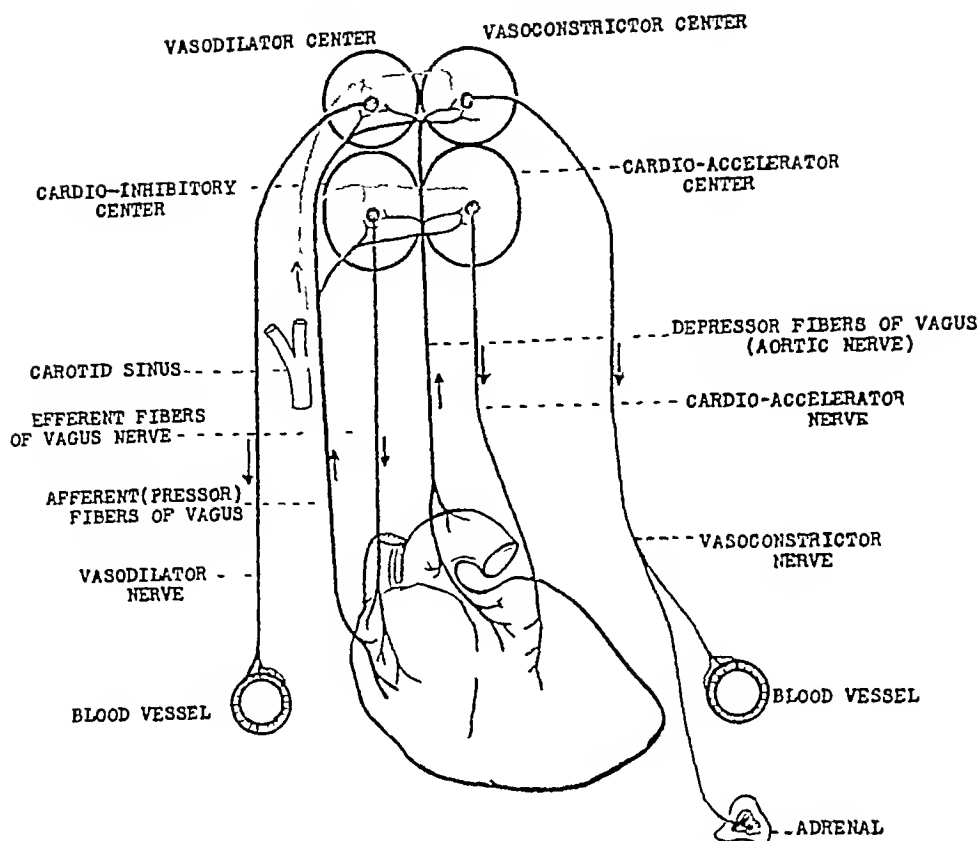


FIG 27 9 Diagrammatic representation of cardiovascular reflex mechanisms. Afferent vagal fibers in blue, sinus nerve in green, efferent fibers to the heart and to the blood vessels in red. The afferent fibers are represented as causing reciprocal effects upon the medullary centers (see also p 287 and fig 27 11)

The carotid sinus in man

It is a very old observation that pressure upon the carotid artery in the human subject may cause slowing of the heart (Parrv, 1799). Many of the examples in the literature of bradycardia due presumably to pressure upon the vagus were in all likelihood due to a sinus reflex. Indeed it is very questionable whether the vagus can be stimulated by pressure through the overlying tissues, for in operations upon the neck, even pinching the nerve with forceps does not stimulate it, and in animals, the vagus is relatively unresponsive to mechanical stimulation.

Since the discovery of the carotid sinus, many observations upon the depressor reaction in man have been reported. In paroxysmal tachycardia the rate may often be restored temporarily to normal by the application of pressure to the sinus. Weiss and his associates have shown, however, that the sensitivity of the sinus reflex varies con-

siderably in different individuals. He found that in about 30 per cent of persons with normal cardiovascular systems, no response whatever could be obtained. In the remainder the cardiac slowing amounted to less than 6 beats per minute and the fall in blood pressure to less than 10 mm Hg.

HYPERSENSITIVITY OF THE SINUS (DEPRESSOR) REFLEX. In hypertension and arteriosclerosis the reflex is often unusually sensitive. In from 70 to 78 per cent of cases of hypertension observed by Weiss and Baker, a fall in arterial pressure of from 10 to 105 mm Hg (average 40 mm) and a reduction in heart rate of from 4 to 20 beats followed pressure upon the sinus. In a corresponding proportion of arteriosclerotic cases a fall of from 10 to 65 mm Hg and an average reduction in heart rate of 16 beats per minute resulted.

These observations lend no support to the idea that a failure of the depressor sinus reflex as a result of structural changes in the sinus wall is

a factor in the production of essential hypertension. As a matter of fact, marked changes in the walls of the sinus may occur with a normal blood pressure, while hypertension may exist though the sinus shows no abnormality.

THE CAROTID SINUS SYNDROME (VASOVAGAL SYNCOPÉ (LEWIS)) Attacks of dizziness and fainting and sometimes convulsive seizures may result from overactivity of the sinus reflex. An attack may occur without known cause, be induced by emotion or follow slight pressure upon the neck. In one case which has been reported, shaving the skin of the neck overlying the sinus or buttoning a tight collar precipitated an attack. During an attack, marked slowing of the heart and a fall in arterial pressure occur. Extrasystoles, delay in A-V conduction, complete heart block or typical Stokes-Adams attacks may occur. In subjects of these paroxysms hypersensitivity of the sinus mechanism can usually be demonstrated between attacks, slight pressure producing a pronounced depressor reaction (see fig 27 10). In some of these cases no structural abnormality of the sinus is evident, the condition, apparently, being purely functional in nature, in others the sinus shows an aneurysmal dilatation, and in others again, a small tumor in the region of the carotid bifurcation has been responsible. In any event denervation of the sinus relieves the condition. After denervation, the blood pressure rises and the heart rate increases for a few hours but soon returns to normal.

Three types of the carotid sinus syndrome are recognized, *vagal*, *depressor* and *cerebral*. In the first mentioned type, cardiac arrest or marked slowing of the heart (heart block) occurs, which can be abolished by the administration of atropine. The depressor type is marked by vasodilatation and a fall in blood pressure, effects which are not affected by atropine but may be abolished by adrenaline. Symptoms referable to the cerebral centers are temporary blindness, staggering gait and giddiness, sudden and temporary weakness of muscles, none of which phenomena are affected by either atropine or adrenaline.

BALANCED AND RECIPROCAL VASCULAR REACTIONS

In the intact animal the height of the blood pressure at any moment, insofar as the nervous control of the peripheral vessels is concerned, is apparently the algebraic sum of the effects of afferent impulses impinging upon the vasomotor centers. Under ordinary circumstances impulses

arising from the carotid sinus and aortic arch play the most prominent rôles, but impulses from skin, muscles and viscera and from higher nervous centers also exert an important influence. That pronounced effects upon the peripheral vessels can be produced by the irradiation of impulses from higher centers is evidenced by such phenomena as blushing, pallor, erection and certain types of syncope (fainting). The vascular changes observed by Drury and Florey in the mucosa of the exteriorized colon of the dog when the animal was excited, and the changes in splenic volume noted by Hargis and Mann and by Barcroft (p 70) are other examples. Even very mild excitation of psychic centers exerts an influence upon the vascular mechanisms. The psychogalvanic reflex (due to changes in the electrical resistance of the skin) has a vascular basis. A reciprocal relationship also exists between splanchnic and cutaneous vascular areas on the one hand and the vessels of the muscles on the other. Adrenaline, for example, causes dilatation of the latter vessels accompanied by vasoconstriction in the skin and abdominal viscera. Stimulation of the wall of a large vein or distension of the duodenum causes reflex constriction of the cutaneous vessels, and stimulation of the skin results in an increase of the blood in the liver and in the renal cortex. Furthermore, the muscular and cutaneous tissues may show opposite vascular reactions. Thus, in the dog, cooling of the body causes constriction of the skin vessels and vasodilatation in the muscles

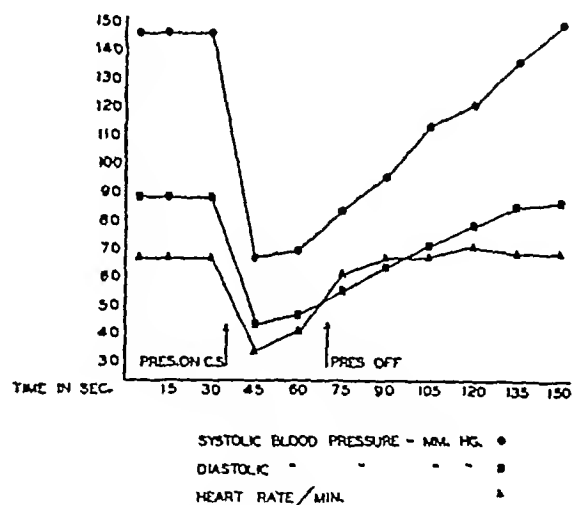


FIG 27 10 The effect on blood pressure and heart rate of pressure on the carotid sinus (hypersensitive) in the human subject. Note that when the pressure is released, the heart rate returns to normal very quickly, the blood pressure more slowly. (After Weiss and Baker.)

An interesting example of the complex nature of the reflex vascular adjustments has been described by Bronk and Gammon. An electrical record obtained from a nerve twig coming from a Pacinian corpuscle in the mesentery during perfusion of the mesenteric vessels showed an increase in the frequency of the afferent impulses when the perfusion pressure was raised. The impulse frequency was reduced by bleeding the animal and increased again when the blood was reintroduced into the body. It is suggested that the Pacinian corpuscles which lie in close relation to the vessels are stimulated when the latter dilate. Messages pass to the vasomotor center which then discharges vasoconstrictor impulses to the vessels of the area from which the afferent impulses arose. Thus, undue distention of the vessels and pooling of blood in the splanchnic region is prevented. It has also been shown by Izquierdo that stimulation of the peripheral end of the splanchnic nerve causes a much greater rise in blood pressure than usual if the carotid sinuses have been excluded from the circulation (by clamping the carotids). This observation indicates that ordinarily the pressor effect of splanchnic stimulation is largely counteracted by a depressor reflex initiated from the sinus. Chemoreceptors have been demonstrated as well in the mesentery of rats and mice by Hollingshead. They give rise to strong vasopressor and weak respiratory responses when stimulated by anoxia (cyanide, or inhalation of nitrogen).

Löwen reflex. This is the term applied to the reaction first described by Löwen, in which a local dilatation of vessels accompanies a general vasocon-

striction. When, for example, the central end of an afferent nerve to an organ is stimulated while its efferent vasomotor fibers remain intact, a rise in general blood pressure occurs together with a dilatation of the vessels of the organ. It is evident that in the intact animal such a mechanism, brought into play through afferent impulses arising within the organ itself, will provide it with an increased blood flow during activity.

The reciprocal action of the medullary centers. The experiments of Bayliss and others indicate that the reflex vasomotor effects involve both centers in a reciprocal manner. In the depressor reflex the tone of the vasodilator center is raised while that of the vasoconstrictor center is lowered. Conversely, the vasodilator center is depressed and the vasoconstrictor center excited in a pressor reflex. The reciprocal reactions are shown diagrammatically in figure 27 11.

It has been generally conceded that in a depressor reflex, loss of vasoconstrictor tone (p 282) is accompanied by increased tone of the vasodilator center. It is also granted, of course, that in a pressor reflex the vasoconstrictor center is stimulated, that depression of vasodilator tone accompanies the constrictor effect has not always been so apparent. The following experiment of Bayliss shows that the latter reaction also occurs. The vessels of the salivary gland were maximally dilated by means of heat and deprived of their constrictors by sectioning the cervical sympathetic. A pressor reflex elicited by stimulating the afferent end of a somatic nerve then resulted in a reduction in blood flow through the gland, which was attributed to a reduction in vasodilator tone.

Vasodilator tone, mediated by fibers running in the posterior spinal nerve roots, has been clearly demonstrated by Bach. He observed that the blood pressure which had been lowered by stimulation of the depressor nerve rose 20 mm. Hg or more above the original (normal) level when, by an ingenious maneuver, all the posterior nerve roots were suddenly and simultaneously severed. The blood pressure was sustained at the higher level but fell again below normal (90 mm Hg) when the sympathetic was excised.

FAINING OR SYNCOPE

Fainting or syncope (G *syncope*, a cutting short, a swoon) is a rather indefinite word which describes a sudden loss of consciousness with partial or complete muscular relaxation. It is a symptom of a large number of pathological states, e.g., carotid sinus hypersensitivity, reflex cardiac inhibition, myocardial disease, hemorrhage, shock, anoxia, heart block (Stokes-Adams' disease), orthostatic

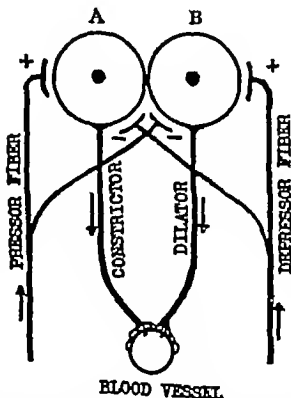


FIG 27 11 Illustrating the reciprocal action of the centers in vascular reflexes. A, vasoconstrictor center, B, vasodilator center

hypotension (ch 16), certain cerebral conditions and even hysteria. The type of fainting or syncope considered in this section is of a more physiological nature, it occurs in perfectly healthy persons, and is of common occurrence. Among the chief causes are emotional disturbances—fear, nervous shock, the withdrawal of blood for transfusion purposes, or even the sight of blood, pain, a sudden rising erect from recumbency, overheating of the body or fatigue. Fainting from such causes is not of cardiac origin. The cardiac output is not necessarily reduced, and when a reduction does occur, it is not great enough to be significant. There is a fall in blood pressure of 25 mm Hg or so, and diminished cerebral blood flow due to peripheral vasodilatation, the heart is slowed. The hepatic blood flow, as determined by the bromsulphalein method and catheterization of the hepatic vein, is reduced, due apparently to vasoconstriction. In this vasodepressor type of fainting, the vasodilatation occurs mainly in the skeletal muscles whose blood flow is increased. The blood pressure in the veins draining muscular areas is raised, and the blood is more arterial in color than ordinarily (a fact first remarked upon by John Hunter). Vasodilatation in the splanchnic area occurs to a lesser degree. The faint is ushered in by pallor, often a feeling of nausea, sweating and a sensation of giddiness. Brun and associates observed that fainting caused by tilting the body into the upright position was accompanied by oliguria and an increase in urinary chloride, an effect which they proposed was due to the release of the antidiuretic hormone of the hypophysis. Later, Taylor and Noble found a pituitary-like principle in the urine of subjects in which fainting had been induced.

Engle advances an interesting hypothesis to account for the muscular vasodilatation in a vasodepressive faint. The effect, he postulates, is a primitive reaction which would necessarily immediately precede and accompany muscular activity whether for the purposes of struggle or flight when some alarming event suddenly occurs, yet at the moment of the event leading to the faint, any actual purposeful movement is voluntarily inhibited or impossible. It will be recalled that the liberation of the antidiuretic principle of the posterior pituitary occurs also in muscular exercise (ch 35).

VASOMOTOR REACTIONS DURING MUSCULAR EXERCISE

Constriction of the splanchnic vessels occurs at the commencement of muscular exercise. This is

attributed to the reception by the medullary centers of impulses discharged from the motor and psychic areas of the cerebral cortex. It is said that splanchnic vasoconstriction actually anticipates the commencement of the muscular effort and may even occur at the thought of taking exercise, before any actual movement is made. If the muscular effort is very strenuous, a rise in hydrogen ion concentration of the blood may occur and act as a stimulus to the vasomotor center. In ordinary exercise, however, no appreciable change in reaction of the blood as a whole occurs and even in the more severe types of muscular exertion a change in blood reaction is of minor importance in the production of the splanchnic vasoconstriction. The small vessels of the muscles dilate during exercise but this, as described on page 289, is a direct action upon the vessels of substances produced in the muscles themselves, though the vasomotor centers may play a part, the evidence is against the possibility. Grant found that vasodilatation occurred in normal and sympathectomized limbs to about the same extent during exercise. The arterial blood pressure is highest in the earlier periods of the exercise when the systolic level may reach a height of 180 mm Hg or more. Vascular readjustments, e.g., cutaneous vasodilatation and possibly a reduction in the degree of splanchnic vasoconstriction, tend to occur as the exercise continues, some reduction in the blood pressure results. Within 10 seconds or so after the cessation of the exercise, the pressure falls to normal but rises abruptly again to its previous level, then gradually declines and finally reaches the normal value in from one to four and a half minutes. These fluctuations in the arterial blood pressure following the exercise are not, however, of nervous origin but are due to mechanical factors (see p 155).

An interesting reflex mechanism in the hypertension of exercise has been demonstrated in the human subject by Alam and Smirk. Contractions of even a small group of muscles, e.g., those moving the little finger, while the circulation through the arm was arrested, caused a rise of as much as 70 mm Hg. Cerebral influences cannot be mainly responsible for the rise since it persisted, provided that the arm circulation was occluded, for some time after the exercise had ceased. Any effect of muscle metabolites upon the nervous centers is, of course, excluded since none could enter the general circulation. For the same reason increased venous return to the heart cannot be a factor in the pressor effect. Arrest of the circulation alone

was ineffective. The phenomenon, like the reflex acceleration of the heart (p 247), is apparently due to the stimulation of afferent endings in the muscles by metabolites accumulated as a result of the circulatory arrest. Of how much importance such a reflex is in ordinary exercise, i.e., with free circulation through the muscles, it is difficult to say

The effects of exercise upon other physiological functions are dealt with in other sections of this book. The reader is referred to the index under Muscular exercise

AXON REFLEXES

These are not true reflexes, since no nerve cell is involved. The efferent and afferent limbs of the axon reflex arc are formed by the branching of a single nerve fiber. A stimulus applied to one branch sets up an impulse which travels centrally to the point of division from where it is reflected down the other branch to an effector organ. The most familiar type of axon reflex is that which involves a sensory nerve fiber and through which vasodilator effects are brought about. Numan Bruce has made a study of the superficial vasodilatation which results from cutaneous or conjunctival stimulation, and offers convincing evidence for the dependence of these reactions upon axon reflexes. Spiess had remarked some years before that in inflammatory states of the superficial tissues the more painful the part became, the greater was the development of heat and redness. The degree of dilatation of the vessels seemed to be dependent upon, and directly proportional to, the pain experienced. A local anesthetic reduced the pain and the inflammatory reaction as well. Bruce found that the dilatation of the vessels which results from the application of an irritant, such as mustard oil, to the conjunctiva of an animal can be prevented if the sensory

nerve endings are first paralyzed by means of cocaine (or better, alvpin, which has no direct effect upon the vessels). On the other hand, if the mustard oil be applied shortly after division of the fifth nerve, which also renders the conjunctiva insensitive, the usual inflammatory reaction occurs. Yet if sufficient time be allowed to elapse after division of the nerve for degeneration of the sectioned sensory fibers to take place, the conjunctival reaction to irritants cannot be elicited. Corresponding results were obtained in experiments upon the skin of the trunk. The employment of a local anesthetic abolished the reaction. It was obtainable after section of the posterior roots peripheral to their ganglia, so long as degeneration of the sensory fibers had not occurred, but was abolished after this. The vasodilator reactions cannot, therefore, be due to a central reflex, yet they are dependent upon the integrity of the sensory nerve fibers. Since no peripheral nerve cells are known to exist through which such a reaction could occur, the latter must, it is concluded, be due to an axon reflex. The impulse evidently passes up a sensory fiber and then down (i.e., antidromically) a collateral branch supplying an arteriole (fig 27 12). See also page 276. Knowledge of these local reactions has obviously wide applicability to the investigation of lesions of the superficial tissues associated with vascular changes, and especially those of an inflammatory nature. They afford another example of the close association between the transmission of pain and vasodilator effects.

It may be mentioned incidentally that axon reflexes are not confined to afferent fibers and to vasodilator reactions. The discovery of axon reflexes was made by Langley and Anderson in the investigation of an observation of Sokolowin's, namely, that when the central end of the hypogastric nerve of one side was stimulated, contraction of the opposite half of the bladder occurred. The impulse was shown to have passed centrally for a short distance to an axon branch, along which it was transmitted to the inferior mesenteric ganglion of the opposite side from where it was relayed to the vesical muscle. Axon reflexes were later shown by these observers to occur in other parts of the autonomic nervous system (in the ganglia of which no true reflexes occur). The preganglionic fibers of a given spinal segment do not necessarily terminate in the first ganglion of the lateral chain which they enter but pass along the chain for some distance giving off collaterals (axon branches) to cells in the ganglia through which they pass, or they may make no connection until they have reached a

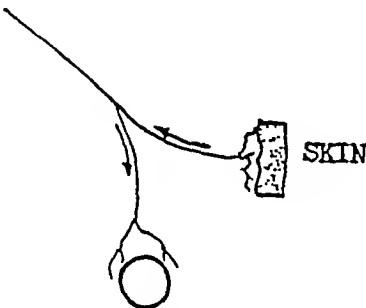


FIG 27 12 Axon reflex. Description in text

level several segments below that of the segment from which they originated. Therefore, when the sympathetic cord was cut and its central end stimulated, pilomotor and vascular effects were produced in skin areas lying at a higher segmental level than the point of stimulation. That axon reflexes of a more local nature occur in the autonomic plexuses of the intestine and other hollow viscera is generally conceded.

THE BLOOD GASES, LACTIC ACID AND HISTAMINE IN THEIR RELATION TO THE VASCULAR MECHANISMS

CARBON DIOXIDE EXCESS AND OXYGEN LACK ASPHYXIA

The vasomotor centers are highly sensitive to the gaseous composition of the blood flowing through their vessels.

A high carbon dioxide tension or a low oxygen tension causes an increase in vasoconstrictor tone and a rise in blood pressure. The persistent elevation of the blood pressure which results from a rise in intracranial pressure (e.g., cerebral tumor or hemorrhage) is due to compression of the medullary vessels and the interference with the blood supply to the centers, and the marked pressor effect induced by asphyxia is the result of the excitation of the vasoconstrictor center by the more venous character of the blood. Three stages in the circulatory effects of asphyxia can be distinguished: (1) For a time after the respiratory muscles of an animal have been paralyzed by curare, little change in the level of the blood pressure occurs. (2) In a minute or two the blood pressure, as a result of the arteriolar constriction, commences to mount and may soon reach a value more than double the normal. The vasoconstriction is undoubtedly enhanced by adrenaline liberation, the hormone exerting a direct action upon the vessels. The heart at this time, as a result of the fuller relaxation of the cardiac muscle and the increased venous inflow, beats forcibly, the rate is slowed in consequence of the rise in blood pressure as well as by the action of carbon dioxide upon the cardio-inhibitory center and the tissue of the sino-auricular node. The effect of the carbon dioxide excess upon auriculo-ventricular conduction may result in heart block (p. 227). The capillaries and small veins are dilated and intense cyanosis (p. 437) occurs. (3) The blood pressure falls. This is due to failure of the heart as a result of the anoxemia, and not to the release of the arterioles from the constrictor

influence, for, if the volume of the kidney is recorded at this time it will be found that no change occurs before death.

The hypertonicity of the vasoconstrictor center in asphyxia might be due either to the carbon dioxide excess or to oxygen lack. Mathison, however, has studied these two influences separately. When an animal breathed an air mixture containing 10 per cent of carbon dioxide with an adequate percentage of oxygen, the arterial blood pressure changed almost immediately, rising within less than a minute to double its previous height, the intestinal volume fell (fig. 27 13, a). Oxygen lack alone, produced by the inhalation of nitrogen, caused a pressure rise of about the same magnitude, the response was delayed for half a minute or so but was then abrupt (fig. 27 13, b). Injections of lactic acid or other organic acids into the blood stream produced effects which were in general similar to those of carbon dioxide excess or oxygen deficiency (fig. 27 13, c). These three factors, carbon dioxide excess, oxygen lack and a rise in hydrogen ion concentration by the injection of lactic acid, act upon the vasomotor center and the chemoreceptors of the carotid and aortic bodies (especially the latter) in a manner analogous to that described for respiratory control (ch. 33). Thus, the action of CO_2 is exerted mainly upon the center, whereas anoxemia (or cyanide which suppresses oxidative processes) and acid injections act in a common fashion by raising the hydrogen ion concentration of the blood bathing the chemoreceptors.

The spinal centers are also sensitive, though to a less degree, to a rise in hydrogen ion concentration. In the decapitate animal, for example, under artificial respiration, a rise in blood pressure does not occur until the carbon dioxide percentage in the inspired air has reached 20 per cent, whereas breathing air containing 5 per cent carbon dioxide may be sufficient to excite the medullary center.

The local effect of carbon dioxide excess or of lactic acid upon the peripheral vessels is one of dilatation. The central and peripheral effects of the hydrogen ion are therefore opposite in direction. During exercise, for example, the acid metabolites formed in the active muscles cause dilatation of their minute vessels. On the other hand, any rise in the hydrogen ion concentration of the blood as a whole which may result will, through an action upon the medullary center, cause splanchnic vasoconstriction. Such an effect will contribute

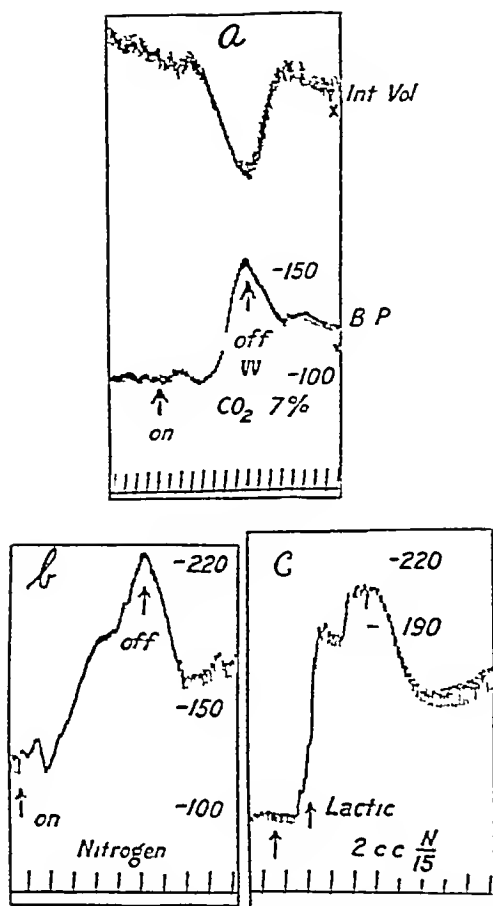


FIG 27.13 Showing the pressure effect of carbon dioxide excess, oxygen lack and lactic acid injections. *a*, tracings of intestinal volume (upper curve) and of blood pressure in an experiment in which a mixture containing 7 per cent CO₂ was breathed. *b*, shows the rise in blood pressure during the breathing of nitrogen, and *c*, the rise in arterial pressure resulting from the injection of lactic acid. (After Matheson.)

to some degree, to the general rise in blood pressure accompanying exercise (p 289). The high arterial pressure, combined with dilatation of the vessels within the muscles themselves leads, of course, to a maximum blood flow through the active tissues. It has already been mentioned, however, that unless the exercise is very strenuous no change in blood reaction occurs, and in any event it appears that, as compared with purely nervous mechanisms (e.g., irradiation of impulses from cerebral centers), the effect upon the vasomotor center of an increase in hydrogen ion concentration plays a minor rôle in the hypertension of muscular effort. The local vasodilator effect

of the hydrogen ion, on the contrary, is of the utmost importance.

The work of Anrep and his associates indicates that histamine liberation is also a factor in the local vasodilator reaction. These observers found that, whereas the histamine concentration in the venous blood coming from a resting muscle was the same as that in the arterial blood, during contraction of the muscle, the histamine concentration in the venous blood exceeded that in the arterial blood.

CARBON DIOXIDE DEFICIT

Carbon dioxide depletion (acapnia) produces effects upon the vascular system which, in general, are the reverse of those resulting from carbon dioxide excess. The tone of the constrictor center is lowered, arteriolar dilatation results and the blood pressure falls. The direct effect of the carbon dioxide lack upon the peripheral vessels is to increase their tone. The capillaries and venules are apparently more responsive to the local action than are the arterioles upon which the central effect predominates.

The effects of carbon dioxide deficit upon the circulation have been studied by Dale and Evans. They found that when cats were over-ventilated at the rate of from 100 to 180 respirations per minute there occurred a rapid and profound fall in blood pressure (to 30 or 40 mm Hg within a minute, or two). In these experiments the depressor effect was due chiefly to the loss of arteriolar tone, the effect of the loss of carbon dioxide or of the respiratory movements (p 250) upon the action of the heart played an unimportant part. When the excessive ventilation was carried out, using an air mixture containing 5 per cent carbon dioxide, the fall in pressure did not occur, or was observed only at the beginning of the period of hyperventilation. The depressor effects of acapnia were also obtained in decapitated animals, thus indicating that the spinal vasomotor centers are also highly sensitive to carbon dioxide deficit. After destruction of the cord, changes in the tension of the gas caused reverse effects, namely, a rise in pressure with carbon dioxide deficit and a fall with carbon dioxide excess, the direct actions upon the vessels (constriction and dilatation respectively) being then unopposed by the central influence.

In the human subject the effect of forced breathing upon the blood pressure varies. In the majority a fall

in pressure results but in some the pressure remains unchanged, in a few it rises. The depressor effect, when it occurs, is not due altogether to the effect of carbon dioxide deficit upon the vasomotor center but is largely the result of the mechanical effect of the forcible expiratory movements (p. 250) which impede the venous return to the heart (Vincent and Thompson). This is evidenced by the fact that in many persons the depressor effect can still be produced when the forced breathing is carried out with air containing a high percentage of carbon dioxide. In those persons in whom the pressure does not fall, its failure to do so may be due to the irradiation of impulses from the motor area to the vasomotor center which enhances the latter's tone, or to the compensatory vasoconstriction resulting from the local effect of the carbon dioxide deficit. Vasoconstriction of the cutaneous vessels is often clearly evident from the pallor which occurs in the period of apnea following forced breathing, the cutaneous circulation is sometimes slowed to such a degree that cyanosis appears. Stewart, by means of his calorimetric method, has also demonstrated under such circumstances the slowing of the blood flow through the hands. Furthermore, if the vessels are kept maximally dilated by means of a hot bath, in a subject who ordinarily either shows no blood pressure effect as a result of forced breathing or gives a pressor response, then acapnia causes a depressor effect. The antagonism between the central and local effects of carbon dioxide upon the vessels can also be demonstrated by applying a tourniquet to the arm. Forced breathing then causes a much less marked effect upon the vessels of the asphyxiated arm (owing to the high carbon dioxide tension in the blood of the obstructed vessels) than upon those of the opposite arm.

The actions upon the vascular system of adrenaline, noradrenaline (ch. 59), pituitrin (ch. 57), and of VEM and VDM (chs. 3 and 27), are dealt with elsewhere.

Traube-Hering waves

These are slow, rhythmical waves which appear on the blood pressure tracing of an animal poisoned by curare, absinthe, morphine and certain other drugs. They may also appear in asphyxia or in any condition in which the oxygen in the blood supplying the medulla is markedly reduced or the carbon dioxide content increased. They may therefore appear in hemorrhage or when (as a result of raised intracranial pressure) the circulation through the medulla is interfered with. The frequency of the waves is from 5 to 10 per minute. They are due, apparently, to periodic variations in tone of the vasomotor center and should not be confused with those of splenic origin (p. 70).

THE CONTROL OF THE VEINS

The veins receive constrictor fibers from the sympathetic. The innervation is not confined to

the minute veins, but includes the larger superficial veins of the limbs, intestines, spleen, liver and kidney. There is no evidence, however, that the venae cavae or the veins of the muscles are furnished with constrictor nerves. Like the arteries, veins exhibit constrictor tone, dilating when their nerves are sectioned. The vein wall responds readily by constriction to direct mechanical stimulation, e.g., puncture by a needle. Certain superficial veins receive dilator impulses (antidromic) via sensory fibers. Dilatation of the inferior mesenteric vein has also been observed following stimulation of the posterior nerve roots of the lower thoracic segments. In emotional states constriction of the superficial veins has been observed and reflex changes in caliber, analogous to those occurring in the arterioles, result from various types of peripheral stimulation. According to Heymans the veins take part in the pressor or depressor reflex evoked from the carotid sinus.

Carbon dioxide acting locally upon the walls of the small veins exerts a dilator effect, reduction in carbon dioxide tension increases venous tone. Yandell Henderson and associates attach a great deal of importance to the tonic influence exerted upon the venous mechanism by carbon dioxide. They claim that acapnia produces constriction of the peripheral veins and so interferes with the return of blood to the heart. The rise in venous pressure, which results from breathing an air mixture rich in carbon dioxide they ascribe to relaxation of the smaller veins and the increased flow of blood from the arterial side.

The reactions of the veins to adrenaline, pituitrin and several drugs are similar to those of the arterioles to such agents.

MEANS EMPLOYED TO DEMONSTRATE THE PRESENCE OF VASOCONSTRICTOR OR VASODILATOR FIBERS IN A PERIPHERAL NERVE TRUNK

If both types of vasomotor fibers exist in a given nerve, the constrictor effects predominate when the nerve trunk is stimulated, and may obscure entirely any coincident response of the dilator fibers. In order, therefore, to demonstrate the presence of the latter it is necessary to remove the constrictor influence before stimulating the nerve. This may be accomplished by sectioning the nerve and allowing time for *degeneration of the vasoconstrictors* to occur, which is some time (2 to 3 days) in advance of the degeneration of the vaso-

dilators At the end of this time, therefore, pure vasodilator effects are obtained The two types of fibers also respond more readily to *different kinds of stimuli*, and this fact may sometimes be employed to detect the presence of vasodilator The vasoconstrictors respond more readily to strong electric shocks of high frequency, the dilators to comparatively weak and slow rhythmic shocks The latter also respond more readily to mechanical stimulation *Ergotoxine* is a drug which acts in a selective manner, as shown by Dale, by abolishing motor effects of the sympathetic nervous system The vasoconstrictors are paralyzed but the vasodilators are unaffected Therefore when a nerve, such as the splanchnic, which contains both types of fibers, is stimulated after administration of this drug, the response is one of pure vasodilatation

Several methods are available for the demonstration of the actual vasoconstrictor or vasodilator effect (1) *Inspection* of the organ or vascular region supplied by the nerve under investigation Flushing or blanching indicates vasodilatation or vasoconstriction, respectively The method is particularly suitable to superficial and transparent structures, such as the frog's web or the ear of the rabbit, the latter was employed by Claude Bernard when he discovered the vasomotor nerves

(2) *Change in temperature* of a superficial part supplied by the stimulated nerve The blood in the deeper tissues is considerably warmer than that flowing through superficial vessels When the arterioles dilate, the part becomes flushed with blood drawn from deeper regions, i.e., the cutaneous blood flow is increased, the temperature rises in consequence A reverse effect upon the superficial temperature is caused by vasoconstriction The temperature changes may be detected by means of a sensitive thermometer held in contact with the surface or, more precisely, by a thermo-electric couple inserted into the tissue, e.g., the skin, mucous membrane, etc

(3) *Plethysmographic method* The principles of this method have been explained elsewhere (p. 179) It is applicable to the demonstration of vascular changes in such organs as the kidney, the intestine or a limb, which can be isolated from surrounding parts The instrument registers changes in volume and these, when they follow stimulation of the nerve supplying the region, are taken to indicate corresponding changes in the caliber of the vessels, and in the quantity of

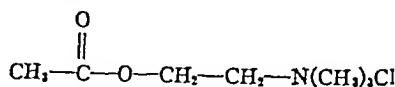
contained blood Both artery and vein leading to and from the organ must be equally free from compression, since interference with the flow through one or other vessel will cause alterations in volume of the part When a limb is under investigation it is usually desirable to eliminate muscular movement by the administration of curare It is essential that changes in volume resulting from passive distensions or collapse of the vessels by changes in the general arterial blood pressure be distinguished from active vasodilatation and vasoconstriction, respectively That is, a rise in the arterial blood pressure may cause greater filling of the vessels in a purely passive manner and swelling of the organ in consequence—a mere distension of the vascular bed On the other hand, a fall in general blood pressure may drain blood from the organ and cause its volume to shrink For these reasons a blood pressure tracing must always accompany the plethysmographic records Only when the blood pressure tracing remains unchanged or the two tracings take opposite directions, i.e., a rise in arterial pressure accompanied by a fall in volume or a fall in pressure coinciding with a rise in volume, can a definite conclusion regarding the change in the caliber of the vessels be drawn (fig. 27.14)

(4) *Outflow from veins* In a small organ, such as the salivary gland, changes in the blood flow through it may be detected by noting the rate of flow and the brightness of color of the blood issuing from the veins Usually the smaller veins are tied off and the blood as it issues from one or two larger veins is collected in a graduated receptacle

The *thermo stromuhr* (p. 175 and p. 327) may also be employed to determine variations in the volume flow through the organ, the measurement being made either of the ingoing or outgoing blood

VASODILATOR AGENTS The results of experimental work in the last few years have added several new names to the list of vasodilator substances which may be extracted from the tissues

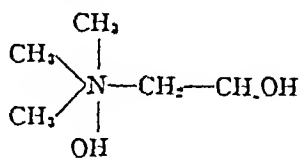
Acetylcholine Much knowledge has been gained in recent years concerning the physiological significance of this choline ester Its formula is



Acetylcholine chloride

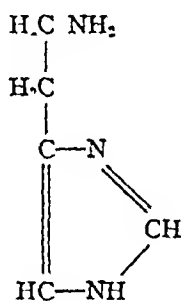
Acetylcholine, as discussed elsewhere (ch. 72), is liberated from parasympathetic postganglionic nerve endings and from the preganglionic endings of both

parasympathetic and sympathetic fibers *Choline* itself, which has the following formula



and as a constituent of lecithin and sphingomyelin (p 689), has an extremely wide distribution in the body, is a well-known vasodilator. Its activity is of a low order in comparison with acetylcholine, the vasodilator effect of choline being increased by at least one thousand times after acetylation. There appears to be very little free choline in the body, and the amount of acetylcholine, rather than choline, appears to vary under physiological conditions. Choline acts in the same manner as acetylcholine on the blood vessels. Its action on liver fat has been discussed elsewhere (p 696).

Histamine



Histamine is also very widely distributed in the body. It can be readily formed *in vitro* by the action of bacteria belonging to the colon typhoid group upon the amino acid histidine. This mechanism probably accounts for the formation of histamine in the intestine. The method by which histamine is formed in tissues has not been established. Small amounts are present in muscle, while the lung in herbivora and the liver and intestinal tract in omnivora may contain relatively large amounts. The concentration of histamine in human blood is from 1 to 8 micrograms per 100 cc. This is contained mainly in the white cells and most probably in the eosinophils (Code). A substance indistinguishable from histamine is apparently responsible for the first part of the triple response in skin (p 315) but chemical identification has not been made. "Gastrin" and histamine (p 509) are indistinguishable. The subcutaneous injection of one-quarter to one-half milligram of histamine produces a copious flow of acid gastric juice in man (p 518). The mechanism of action of histamine on the blood vessels varies with the species studied, but in man both arterioles and capillaries are dilated. The intravenous administration of a ten-thousandth of a milligram causes an appreciable fall of blood pressure in the etherized cat.

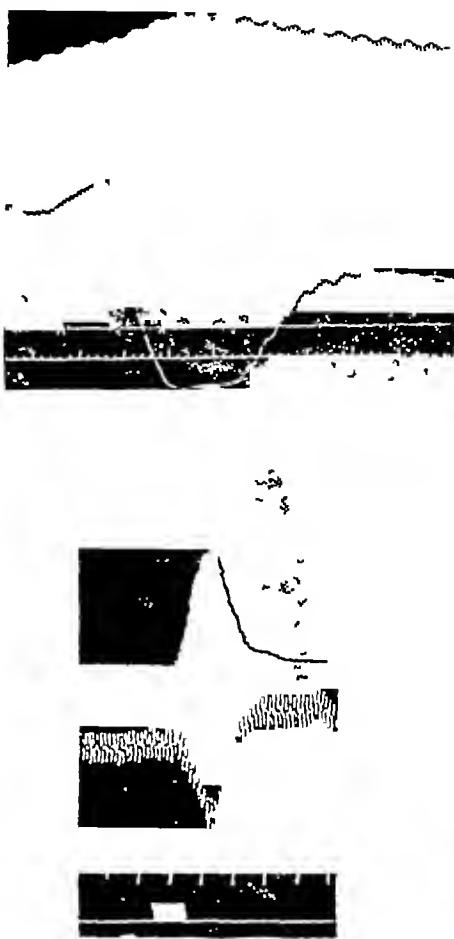


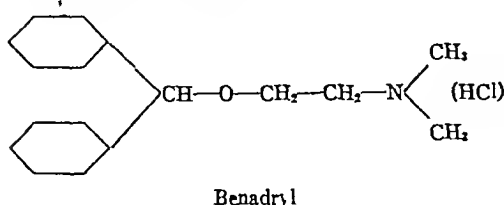
FIG 27 14 Upper figure shows simultaneous tracings of carotid blood pressure and volume of kidney. Between X and X the peripheral end of the divided tenth dorsal nerve was stimulated. Time-marking = seconds (After Bradford.) Lower figure shows the effect upon the kidney volume (PI) and arterial pressure (BP) of stimulating the central end of the vagus, T, time-marker, S, stimulus (After Bayliss.)

One-thousandth of a milligram given intravenously produces a sharp fall of blood pressure in unanesthetized man. The unanesthetized organism is not so susceptible to histamine as the anesthetized. While Abel's researches indicated the wide distribution of histamine in biological material, the base was first prepared in crystalline form from perfectly fresh tissue by Best, Dale, Dudley and Thorpe. The possibility that histamine may possess physiological significance has been materially increased by the finding that it is liberated in anaphylactic shock from the liver in the case of the dog and from the lung in the guinea-pig. Histamine may not be the only substance liberated in this condition but it accounts for most of the signs of anaphylaxis. It is important also, as Kalk has shown, that irritation of the skin in certain susceptible people (dermographism) produces a definite secretion of acid gastric juice quite comparable to that which the sub-

cutaneous injection of histamine might evoke. Histamine is destroyed by an enzyme system, histaminase (Best and McHenry), which is found in various tissues, but particularly the kidney and small intestine. The physiological significance of this enzyme has not been established.

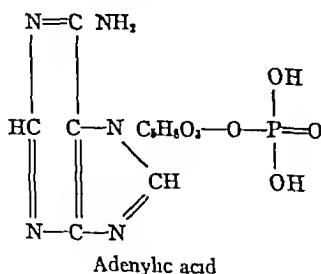
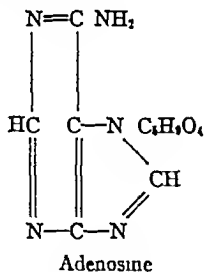
A number of antihistamine agents, that is, substances which antagonize the action of histamine, e.g., benadryl, pynbenzamine, and neoantergan (2786 RP) have been synthesized.

The formula of benadryl is —



These substances are employed in various allergic conditions. They have a protective action against anaphylactic shock in the guinea pig, and against the action of histamine in this species, the latter action is employed as a test of their efficacy.

Adenosine and derivatives



In 1929 Drury and Szent-Györgyi showed that the active constituent of certain extracts from skeletal and cardiac muscle which produced a characteristic condition of heart block in the guinea pig was adenosine. This substance was also shown to have a peripheral vasodilator effect, which has been found to be exerted mainly, if not entirely, upon the arterioles. The heart block (p. 227) produced by adenosine is relieved by barium. Adenosine administered subcu-

taneously causes a migration of leukocytes to the site of injection, an effect which distinguishes it from histamine. Histamine and adenosine may both be concerned in inflammatory reactions and in the effects upon the circulation produced by extensive tissue damage. Their exact significance in this connection awaits further investigation. Adenosine and adenylic acid are relatively inert vasodilators with a potency of somewhat the same order as choline. They may be present, however, in considerable amounts in certain tissues.

Kallikrein Kallikrein is the name given by Frey and Kraut to a vasodilator substance obtained from normal urine. This substance has not yet been obtained in pure form, but it probably differs from all other vasodilators. It does not constrict the bronchioles in the guinea pig, as histamine does, it lowers the blood pressure of the rabbit, while histamine does not do so when the ordinary anesthetics are used. It is rapidly destroyed by blood serum, but the activity is recovered when acid is added. It is suggested that kallikrein is liberated when the tissues become acid through metabolic activity. This would, of course, produce vasodilation and increased blood flow. The site of formation of kallikrein is not known. It is present in relatively high concentration in the pancreas from which it is most conveniently prepared.

Identification of vasodilator substances In addition to these substances there are undoubtedly others which cannot be exactly characterized as yet. The vasodilators we have mentioned may be identified by the following means. Kallikrein and adenosine and its derivatives are destroyed by acid hydrolysis, choline and histamine are resistant to this procedure. Acetylcholine is very susceptible to alkali and may be completely destroyed in a short time in a solution at pH 8 to 9. Atropine antagonizes the action of choline or of acetylcholine but exerts no effect upon the action of histamine. Positive evidence for the presence of adenosine is its characteristic effect upon the guinea pig's heart. Kallikrein will produce a fall in blood pressure in the atropinized rabbit, which neither choline nor acetylcholine will do, while it exerts no effect on the guinea pig's heart. These vasodilators may also be separated by the use of precipitating agents and by other chemical procedures.

Tetraethylammonium chloride or bromide possesses a pronounced vasodilator action, due to its paralytic action upon sympathetic ganglia and thereby blocking tonic vasoconstrictor impulses. It is employed to distinguish spastic vasoconstriction from structural narrowing of the vessels in hypertension and peripheral vascular disease. The methonium compounds (hexamethonium and decamethonium bromide) are also employed for their action in blocking vasoconstrictor impulses.

VASOCONSTRICTOR SUBSTANCES Among the best known vasoconstrictor agents are *pituitrin* and *adren-*

aline (ch 57, 59) and compounds chemically related to the latter, e g, *ephedrine*, *tyramine* and *benzedrine*. *Barium chloride* is also a powerful vasoconstrictor. *Tobacco smoking* is generally stated to cause peripheral vasoconstriction, the effect being usually not demonstrable unless the smoke is inhaled. The experiments of Bolton and associates and of Mulinos and Shulman throw doubt upon the conclusion that the vascular effect is due to the absorption of some constituent of the inhaled smoke. They found that a deep inspiration of pure air caused a marked reflex constriction of the cutaneous vessels of the forearm and hand. The phenomenon is enhanced by an irritant or painful stimulus, e g, inhalation of smelling salts, pinching the skin, etc., and is more pronounced if the respiratory movement is mainly thoracic. Helmer and his associates have shown that the vasoconstrictor agent found in the urine of smokers is nicotine, which they have isolated in crystalline form as a picrate. *Sodium nitrite*, contrary to the general belief, causes, as shown by Weiss and his associates, arteriolar constriction. The fall in blood pressure which follows the administration of an appropriate dose, and which may result in collapse and syncope when the subject stands, is due to the reduction in venous tone and, as a consequence, the pooling of blood in peripheral veins.

PERIPHERAL VASCULAR DISEASE

RAYNAUD'S PHENOMENON AND RAYNAUD'S DISEASE

A distinction must be drawn between *Raynaud's phenomenon* and *Raynaud's disease*. Raynaud's phenomenon is understood to be the intermittent spasm of the small arteries of the extremities of the size of the digital arteries. The affected members, the fingers and less commonly the toes, become cold, numb, waxy pale (dead fingers) or cyanotic. The attacks are precipitated by exposure to cold. It is secondary to some other abnormal state, e g, thromboangitis obliterans, arteriosclerosis, cervical rib, etc.

Raynaud's disease is the occurrence of Raynaud's phenomenon as a primary affection, i e, not associated with some other arterial condition. It is a simple vascular spasm of the digital arteries. It is bilaterally symmetrical. The disease, which is rare as compared with Raynaud's phenomenon, was first described by Raynaud in 1862. The fingers, sometimes the toes, or rarely the ears and nose, are the seat of periodic attacks of vascular spasm. The attack lasts as a rule for a few minutes, but may persist for an hour or two. The color change commences in the finger tips and spreads toward the bases of the fingers. As the attack passes off the part again becomes

cyanotic, then red and hot, the numbness is replaced by burning pain. The pulse at the wrist or ankle persists during the attack. The local asphyxia may lead to ulcers, scleroderma, rarefaction of the terminal phalanges, and other trophic changes. Gangrene sometimes results.

The disease, since Raynaud's first description of it, has been generally attributed to hyperactivity of the vasomotor (vasoconstrictor) nerves. Sympathetic ganglionectomy is, therefore, frequently resorted to in an effort to abolish the attacks. Lewis has shown, however, that a typical attack may be induced in a subject upon whom this operation has been performed, by exposing the affected part to cold (as by immersing it in cold water). He concludes therefore that the fault lies not in the nervous control but in the vascular wall itself. The following observations cited by Lewis substantiate his view.

(1) If one finger of a subject of the disease be immersed in cold water, an attack confined to this finger may be induced. Such a localized result cannot be explained upon the basis of a nervous reflex.

(2) Anesthetization of the ulnar nerve of a normal person by means of novocaine causes dilatation of the vessels of the little finger (removal of vasoconstrictor tone). The vascular spasm in Raynaud's disease cannot be released in this way.

(3) If a subject suffering from Raynaud's disease affecting both hands, and upon whom a unilateral ganglionectomy had been performed, be seated in a cool room with both hands placed in cold water, the vascular spasm which results is more pronounced on the non-sympathectomized side. If, however, the rest of the body is warmed while the hands are immersed in cold water, the attack is more pronounced on the sympathectomized side. In the first experiment the greater degree of spasm on the non-operated side is attributed to the added effect of a vasoconstrictor reflex. The greater degree of spasm on the operated side in the second experiment is attributed to the absence of dilator sympathetic fibers. Though a nervous influence is evident in these observations the essentially local nature of the fault is also indicated.

Ganglionectomy, even though it does not remove the fundamental cause of the condition, does, nevertheless, exert a decidedly beneficial effect. The attacks are less frequent and intense after the operation, normal vasoconstrictor tone and, as just indicated, the reflex responses to cold having been abolished, a more intense reaction of the arterial wall itself must occur before arrest of the circulation to the part can result.

Simpson, Brown and Adson on the other hand, do not consider that Raynaud's disease is due primarily to a fault of the vascular tissue. These observers maintain that only in the advanced stages of the disease is the vascular wall itself abnormal and that in milder cases the fault is essentially vasomotor in character. They point out that Lewis' crucial experiments were performed upon severe or complicated cases of the disease.

ERYTHROMELALGIA

Erythromelalgia is a rare but interesting condition characterized by attacks of painful redness of one or both feet, or occasionally of the hands. The pain is burning in character and is induced by warming or exercising the part, or by allowing it to hang down. Rest, elevation of the part or the application of cold tends to relieve the pain. Erythromelalgia has been attributed to vasodilatation resulting from some abnormality of the vascular nerves. Lewis finds, however, that the essential abnormality in these cases is not vasodilatation, for an equivalent degree of vasodilatation may occur in normal subjects in response to warmth or exercise, yet pain does not result. Erythromelalgia, or *erythralgia* as Lewis prefers to call the condition is not of vasomotor origin. The abnormality in these cases is apparently a hypersensitive state of the cutaneous pain fibers to heat or tension. This "susceptible state" of the skin in erythralgia is altogether analogous to that seen in inflammation, and to that which can be induced in any normal person by certain types of cutaneous injury—exposure to ultraviolet light, repeated rubbing or stretching burns, etc. It is well known that the pain endings of skin injured in these ways are very sensitive to warmth or to tension. Warming the part either by increasing the blood flow through its vessels or by the application of heat causes burning pain. Pain also results when the part is dependent, the engorged vessels then causing tension upon the hypersensitive nerve endings.

Lewis suggests that in the pain associated with erythralgia and with the types of cutaneous injury just mentioned, a chemical substance liberated in the skin serves as the immediate stimulus to the nerve endings. The observation that the pain which follows repeated rubbing or stretching of the normal or of the erythralgic skin is prolonged and intensified by arresting the circulation to the part, supports this conception (p 300).

ACROCYANOSIS

In this disorder the hands and less commonly the feet are persistently cold, blue and sweaty. Exposure to cold intensifies the cyanotic color. In the case of the hands the cyanosis commences at about the level of

the wrist and deepens as it is traced toward the fingers. There is puffiness of the fingers but trophic disturbances are unusual. The milder forms of the disorder are closely allied, according to Lewis, to chilblains. The disorder is due to increased tone of the cutaneous *arterioles* resulting from hypersensitivity to cold. The condition does not have a nervous basis, the fault is in the vascular wall itself, the cyanosis persisting unchanged after anesthetizing the ulnar nerve with novocaine. We have seen that in Raynaud's disease the spasm is of the digital *arteries*. In acrocyanosis the cutaneous circulation is slowed as a result of arteriolar constriction, the partial asphyxia causes capillary dilatation and an increase in the quantity of blood in the skin. The slower blood flow, by allowing the hemoglobin to give up a greater part of its oxygen store, is responsible for the blue tint of the skin, the depth or intensity of the color is due to the fullness of the vessels (p 437).

THROMBOANGITIS OBLITERANS (BUERGER'S DISEASE)

This is an *organic* vascular disease involving, as a rule, the medium and small arteries (and to a less extent the veins) of the extremities. The condition in the majority of cases is confined to the lower limbs. The vessels are stiffened and hard. The adventitia is thickened, the media shows atrophy of its muscle and an increase in connective tissue, and, active proliferation of the intima occurs, several layers of cells being formed. The marked narrowing of the vascular lumen which results is followed by thrombosis. This and not the intimal proliferation itself is responsible for the final obliteration of the vessel. Organization of the thrombus, i.e., its invasion by fibroblasts and its conversion into fibrous tissue, follows. Some restoration of the circulation through the vessel may occur later as a result of the formation of new channels within the substance of the organized thrombus, but whether or not such a process of revascularization results, the blood supply to the part is always greatly reduced.

Among some of the earlier manifestations of the condition are fatigue of the limbs upon exertion, intermittent claudication, hypersensitivity of the vasoconstrictor reactions of the extremities to cold resulting in attacks of pallor or cyanosis, coldness and numbness or a dull ache. A definite reduction in blood flow through the part may be demonstrated by the calorimetric method of Stewart (p 179). Extreme variations in the color of the limb result from altering its position in relation to the level of the heart, when the affected member is dependent, undue redness or cyanosis results, whereas when raised above heart level it becomes intensely pale and waxy in appearance.

As the pathological changes progress, the pulse disappears from the wrist or ankle, or even from the popliteal or brachial artery, ulcers and other trophic disorders appear and ultimately gangrene of the toes or fingers sets in, requiring amputation. The vascular obliteration tends, however, to creep upwards necessitating amputation at successively higher levels.

The cause of the disease is unknown, some believe it to be of infective origin. Excessive use of tobacco is strongly suspected of being a predisposing factor. Buerger and others have remarked upon the very high incidence of the condition in the Jewish race, of a series of 150 cases reported by Brown and Allen from the Mayo Clinic over 50 per cent were Hebrews. The disease occurs almost exclusively in males, whereas Raynaud's disease with which it is likely to be confused, especially in its early stages, affects females predominantly. Another feature distinguishing it from Raynaud's disease is that in the latter, the color of the skin is affected little or not at all by elevation of the limb, and the pulse in the larger arteries does not disappear. It is possible that spasm of the vessels as a result of hyperactivity of the vasomotor nerves is the initial abnormality in thromboangitis obliterans. The constriction of the vasa vasorum, it is conceived, may, by interfering with the nutrition of the arterial wall, initiate the characteristic structural changes.

It is, of course, only during the earlier stages of the condition, i.e., when spasm due to increased vasoconstrictor reactivity is a contributing factor, and before organic changes have progressed to the point where they have occluded the vessels, that treatment of any sort can be expected to bring about any real benefit. When a spastic factor can be demonstrated, sympathectomy frequently results in very notable improvement. Even when the larger arterial vessels are obliterated, sympathectomy, by removing the vasoconstrictor tone of collateral vessels, may be followed by a definite improvement in the blood supply to the part.

There are several methods to choose from for the detection of vascular spasm. The temperature of the part may be taken by means of a skin thermometer or a thermocouple before and after one or other of the following procedures, which, normally, cause vasodilatation and raise the temperature of the part: (a) Heating the entire body, with the exception of the head and the affected part itself, in a cabinet (fig 27 15). (b) The induction of fever by the injection of a foreign protein, typhoid vaccine. (c) Spinal anesthesia which temporarily paralyzes the vasoconstrictors. (d) Anesthetization of a peripheral nerve. (e) The intravenous injection of hypertonic saline by increasing the blood volume will tend to cause a compensatory vasodilatation. A rise in temperature of the part, following one of these procedures, indicates the previous existence of a spastic element. The greater the degree of spasm the more pronounced and rapid is the temperature rise. If the occlusion is entirely organic in nature no change in temperature results.

Instead of recording the skin temperature, the blood flow through the affected part may be estimated by the calorimetric method before and after one of the procedures just enumerated.

Cervical rib may cause changes in the circulation of the upper extremity not unlike those due to thromboangitis obliterans. The subclavian artery tends to be compressed between the bone and the scalenus anticus muscle with consequent reduction in the volume of the pulse of the affected side. But the peripheral vascular disturbances are due mainly, according to the view of Stopford and Telford, to the pressure of the supernumerary rib upon the vasoconstrictor fibers in the lowest trunk of the brachial plexus. Sensory and motor phenomena as a result of pressure upon the somatic fibers also sometimes occur. Constriction of the vasa vasorum with consequent interference with the blood supply of the vascular wall has been suggested as responsible for the structural changes which ultimately lead to thrombosis and occlusion of the arterial lumina. Lewis and Pickering, however, dissent from this view. "Why," it is asked "if pressure upon nerve fibers is the cause of the vascular condition, are the effects always of an irritative nature (vasoconstriction)?" "Why do not paralytic manifestations (removal of vasomotor tone with consequent vasodilation) appear at some later stage?" They suggest that the blocking of peripheral vessels is of embolic origin due to pressure injury of the subclavian artery and the formation upon its wall of thrombi which are whisked away in the blood stream from time to time.

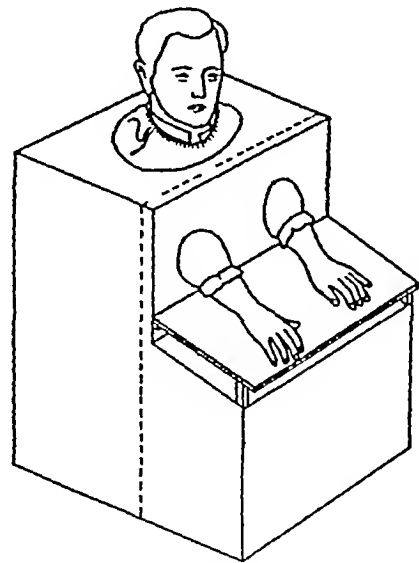


FIG 27 15 Warm chamber with subject in position for observing effects in the hands. The neck and forearms pass through canvas sleeves terminating in purse strings. The chamber is divided into two parts at the dotted line. For use with the lower extremities the front wall is made with simple upright front with two appropriately cut outlets for the thighs. (After Lewis and Pickering.)

INTERMITTENT CLAUDICATION

(*claudicare*, to limp)

This is a condition (described by Charcot in 1856) in which, as a result of organic narrowing of the arteries of a limb and consequent restriction of its blood supply, severe pain is experienced in the muscles during exercise. The pain has been attributed to a muscular cramp or to spasm of the vessels. Neither of these explanations can be entertained, for the muscles are flaccid during the attack. The abnormal stiffness of the arteries seems to preclude the possibility of their being narrowed appreciably by spasm, and the smaller vessels are in all likelihood dilated rather than constricted during the attack.

The essential cause of intermittent claudication is a relative anoxia of the muscles—they are called upon to perform work for which the oxygen supply is inadequate. Lewis has shown that pain identical in character with that occurring in this condition can be induced by exercising any normal limb during the arrest of its circulation. When the circulation is restored, an immediate increase in the volume of the limb occurs which is taken to indicate that the vessels were dilated during the pain. In a patient suffering from the disease in one limb it was shown that the pain occurring in the limb during exercise was practically the same with regard to its time of onset, development and duration as that induced by exercising the sound limb during circulatory arrest.

The fundamental importance of anoxia in the production of the pain is also evident from the observations of Pickering and Wayne who found that exercising the muscles of an anemic subject, in whom there was no evidence of arterial disease, caused the characteristic pain of intermittent claudication. Kissin also showed that exercise performed by normal persons during anoxia (induced by breathing an air mixture containing a low percentage of oxygen) caused the typical cramp like pain. The clinical experiments of Lewis indicate that the direct cause of the pain is not oxygen lack itself but the stimulation of sensory nerves by the metabolic products of muscular activity. Ordinarily these are removed by oxidation, but they accumulate when the blood supply is inadequate. He refers to the pain stimulus as "factor P". The evidence supporting this conception is as follows:

(1) The pain does not vary with the individual contractions but is a steady ache.

(2) Using a standard test (maximal grip exerted by thumb and index finger, recorded isometrically, and repeated at the rate of one per second) it was found that in normal subjects with the circulation to the arm arrested, the pain commenced in about 35 seconds after the commencement of the exercise and took another 53 seconds to reach the point where it became intolerable. The pain disappears within 3 seconds after restoring the circulation—presumably as a result of the removal of "factor P". If, on the other hand, occlusion of the vessels is maintained, the pain persists.

(3) Lewis found that the time of onset of the pain is determined by the total amount of work performed rather than by the length of the exercise period. Thus, when the circulation of a normal limb is arrested, pain ensues after the same number of contractions of equal strength whether they are repeated in rapid or in slow succession. On the other hand, if a constant rate is maintained, the pain follows sooner with strong than with weak contractions. If, however, the circulation to the part is only partially obstructed (Katz, Lindner and Landt) or if air containing a low percentage of oxygen is breathed (Kissin), the amount of work necessary to cause pain is lessened by increasing the rate at which the contractions are repeated—the chemical factor presumably accumulating more rapidly as a result of the shorter time intervals allowed for its removal.

(4) If, after the pain has been relieved by restoring the circulation to the part, the vessels are again occluded and the limb exercised, the time of onset of the pain ensues earlier, the shorter the period during which the blood had been permitted to flow. This result suggests that products accumulated during the previous exercise period, if not given sufficient time to be removed, are carried over to the second period and the concentration necessary for stimulation is reached sooner.

The nature of the chemical pain factor is unknown, Katz and associates believe it to be acid in character and non volatile. The ingestion of sodium bicarbonate was found to increase the amount of exercise required to cause pain.

The pain in angina pectoris is generally believed to be produced through a mechanism essentially the same as that causing pain in intermittent claudication.

"IMMERSION FOOT", "TRENCH FOOT"

During the Second World War this name was given to a state of the vessels and tissues of the feet brought on by severe and prolonged chilling of the feet by cold water. It was seen most frequently in persons who had been shipwrecked and immersed for hours in the sea at northern

latitudes But it may also result if the feet alone are kept cold and moist, sailors working under the stress of wartime conditions in wet boots, or soldiers doing duty in wet trenches (trench foot) may then suffer In cases of immersion foot the feet during the period of exposure are swollen, numb and pulseless Their color varies with the temperature from bright red to deep blue or waxy white, or may be mottled with areas of blue and white or blue and red Within a few hours (2-48) after removal from the water, the feet become hyperemic and severely painful, the pain often being described as burning, scalding or stabbing in character The swelling increases, blistering, ulceration, local wasting of muscles and, in the worst cases, gangrene result Capillary damage permits fluid of high protein content to leak into the tissues These effects are likely to be most severe if the temperature of the members has been quickly restored to normal The hyperemic stage, after from 6 to 10 weeks, merges into one in which the feet are pale and cold and very sensitive to exposure to cold

The hands may be affected by cold water in the same way as the foot but much less commonly, since they are rarely exposed to conditions of such extreme severity

While the members are actually exposed to the cold water, the vessels are in strong constrictor spasm, the metabolism of the tissues both as a result of the cold and the anoxia is greatly reduced, at the lowest temperatures (short of freezing), which may sometimes exist during immersion, metabolic processes may cease entirely It is believed, therefore, that a minor degree of tissue damage occurs at this time and the greater part after removal from the water, when the temperature of the parts rises The vasoconstriction is a reflex response due to the action of cold on the general body surface, as well as to its local effect upon the vascular walls, and also probably to the absence of vasodilator metabolites In the hyperemic stage there is maximum dilatation of the vessels, the warming of the body having overcome the vascular spasm The usual vasodilating metabolites are formed and, in addition, according to Lewis, H-substance is liberated as in reactive hyperemia (p 316) following circulatory arrest from any cause

Since the intensity of the hyperemia and the tissue damage are enhanced by a rise in temperature, every effort is made to prevent rapid warming of the affected members Ice bags or blasts of cold

air are, therefore, applied locally to maintain the skin temperature of the feet around 21°C It is also necessary to keep the entire body at as low a temperature as is consistent with comfort, this is achieved preferably by means of some form of cool cabinet

Frost bite Prolonged exposure to severe cold causes vasoconstrictor spasm, and when freezing occurs the circulation in the small vessels of the part becomes completely arrested After the part has thawed, intense hyperemia usually follows, and most of the tissue damage is believed to be caused at this time The increased capillary permeability due to the freezing together with the high capillary pressure of the hyperemic stage causes edema and swelling of the part This of itself may cause strangulation of the circulation, but thrombosis also occurs If the resulting ischemia is extreme, gangrene follows Measures are directed toward the reduction of these effects by thawing the part slowly and by the administration of an anticoagulant, e g, dicoumarol or heparin

SURGICAL OR WOUND SHOCK—CLINICAL AND PHYSIOLOGICAL CONSIDERATIONS

That state of collapse which follows two, three or four hours after a severe tissue injury is variously spoken of as *surgical, wound, traumatic*, or *secondary shock* It is a condition quite distinct from the so-called *primary shock* which supervenes immediately after an injury *Primary shock* has a nervous basis and is, therefore, also called *neurogenic shock*, pain and psychic factors, through their effect upon the vascular system, play a prominent rôle

Secondary, or surgical shock, is characterized by a profound fall in blood pressure, pallor and coldness of the skin, cyanosis of the finger tips and lobes of the ears, fall in body temperature and of the metabolic rate, rapid shallow breathing, small, rapid pulse, apathy, and other manifestations of collapse Further investigation of a patient or animal in shock will usually reveal marked reduction in the circulating blood volume, even though no hemorrhage has occurred In very severe cases the systolic blood pressure may be no more than 60 mm Hg Oliguria is usual and anuria may occur The renal blood flow, the oxygen consumption of the kidney and the glomerular filtration rate are reduced There is a rise in non-protein nitrogen of the blood, and a fall in the alkali reserve

When the chest of an animal in deep shock is opened, the heart is found to be beating vigorously. The cause of the low blood pressure is, therefore, not primarily cardiac. The heart cavities, however, are incompletely filled. The great veins are depleted of blood, the venous pressure is low, the cardiac output reduced (sometimes to only 10 per cent of the normal) and the circulation through the peripheral vessels slowed.

Reduction in the cardiac output and slowing of the peripheral blood flow occur early in shock and precede the fall in blood pressure

There is constriction of the arterioles of the splanchnic and cutaneous areas. This is probably, in part at least, a compensatory reaction—a response of pressor mechanisms (e.g., sinus and aortic nerves, p. 282, 283) to the underfilled state of the vascular system—but having detrimental effects. Fatal shock has been produced by prolonged vasoconstriction induced by the continuous administration of adrenaline (Erlanger and Gasser).

Inasmuch as many observers have reported an increase in the corpuscular concentration of the blood, or *hemoconcentration* as this has come to be called, it has been the general belief that an essential and fundamental feature of secondary shock is the leakage of plasma from the capillaries as a result of an undue permeability of their walls. But the loss of plasma through excessively permeable vessels *remote* from the injured region has yet to be proved. In animal investigations into secondary shock (as produced by mechanical trauma) hemoconcentration to a degree which would indicate any great reduction in blood volume due to plasma loss is by no means a general finding. Furthermore, no direct evidence of plasma loss into the tissues generally has been secured, and even when hemoconcentration exists there is some indication that plasma is removed from the general circulation by becoming "locked up" or pocketed in the minute peripheral vessels rather than by a general leakage from the vascular system.

Widespread capillary dilatation is a factor which has been generally thought to be present in and largely responsible for the shock state. A considerable body of evidence can be cited in support of this belief. Morrison and Hooker observed an increase in the volume of an intestinal loop when shock supervened, and Mann found that, whereas about 24 per cent of the blood volume of a normal animal is held in peripheral vascular areas and the remaining 76 per cent in

the heart and larger vessels, in shock (produced by exposure and handling of the abdominal viscera) 61 per cent was contained in the peripheral vessels.

Knisely and Block have described a sludge like massing of the blood cells in human capillaries as a result of even mild trauma. Such a phenomenon must be an important factor in reducing the circulating blood volume and slowing the peripheral blood flow, and may even be concerned in initiating shock.

Pooling of blood in peripheral vascular areas could quite conceivably account for the main features of shock. Any great enlargement of the capillary bed would alone cause a fall in blood pressure without any loss of blood from the vascular system. The maintenance of the normal arterial blood pressure (ch. 15) depends upon a nice adjustment of the capacity of the vascular system to the volume of blood. Variation in either factor would have very much the same effect upon the blood pressure. Moreover, pooling of blood in peripheral vascular areas would, by reducing the venous return, deprive the heart of an adequate volume of blood to maintain the normal arterial pressure.

THEORIES CONCERNING THE PRIMARY CAUSE OF WOUND SHOCK

Many theories have been advanced in attempts to explain the genesis of the shock state. Among those which have found favor from time to time with different groups of workers are the following: exhaustion or paralysis of the vasomotor center as a result of its bombardment by sensory impulses from the traumatized tissues (Crile), "blowing off" of carbon dioxide (acapnia) due to stimulation of the respiratory center by afferent impulses arising in the injured part, or during the hyperpnea of the early stages of general anesthesia (Y. Henderson), multiple pulmonary emboli caused by the liberation of fat globules from damaged bone and subcutaneous tissues (Porter), adrenal exhaustion (Swingle), etc. None of these theories has stood the test of critical investigation. There is no evidence that the activity of the vasomotor center is depressed in shock. It has been mentioned that arteriolar constriction rather than dilatation occurs, and Porter found that the vasomotor reflexes, both pressor and depressor, were unimpaired in the shocked animal. Nor is acapnia or alkalosis present in traumatic shock. Acidosis is more likely to appear, though not in a causative rôle, the reduced circulation rate and consequent anoxia of the tissues resulting in the accumulation of fixed acids. Porter's fat embolism theory

has not been substantiated. Swingle and his associates have directed attention to the very close similarity between the effects of adrenal (cortical) excision and shock (ch. 59). Low blood pressure due to plasma loss and collapse are prominent features following suprarenalectomy. The effects of this operation are, however, much slower in onset (several days) and there is no direct evidence that deficiency of the hormone of the adrenal cortex is a factor in the great majority of cases of traumatic shock, and cortical principles are of uncertain benefit (see below). With regard to the adrenal medulla, in Cannon's view this part of the gland is more likely to be overactive than deficient in shock (p. 305).

The toxic theory

As a result of the investigations of a number of workers into the shock state, as seen in wounded soldiers in the War of 1914-18, or as produced experimentally in animals, the theory arose that it was due to an endogenous chemical substance having a histamine-like action or possibly to histamine itself which was formed in or released from the injured tissues, especially muscle. A few years before the first World War, Dale and Laidlaw had reported upon the action of *histamine*. Struck by the resemblance of the effects of this amine to certain features of traumatic shock, Dale, Laidlaw and Richards in 1919 re-investigated its actions. When injected into animals in doses of 1 mg. per kilogram of body weight, histamine exerts a profound depressor effect upon the circulation, namely, a fall in blood pressure, due to peripheral vascular dilatation. Increased capillary permeability with consequent transudation of plasma from the vessels is also a well recognized action of histamine (p. 295). The fall in blood pressure is, of course, the outstanding feature of shock. Vasodilatation and increased capillary permeability leading to reduced plasma volume were (upon the assumption that the blood pressure fall was due to histamine or a histamine-like substance) supposed to exist as well. The reports of hemoconcentration and a reduction in blood volume in shock seemed to confirm these suppositions. Certain clinical observations were also in accord with a toxic theory, a tourniquet applied to a damaged limb, for example, appeared to postpone the development of shock, presumably because a toxic substance was prevented from entering the general circulation. It was also generally conceded that amputation of a contused and lacerated part favored recovery from shock, arrested its further development or tended to prevent its onset. Moreover, injury to muscle, which it was presumed was relatively rich in vasodilator substances such as histamine, was especially likely to be followed by shock.

Thus, the toxic theory of shock in which a histamine-like substance, or histamine itself, was the responsible agent, seemed well authenticated and seemed to be finally established by the experiments of Cannon

and Bayliss. These observers crushed the limbs of animals and reported that a fall in blood pressure did not result so long as the vessels of the injured part were occluded, though the nerve trunks remained intact. Upon restoring the circulation a fall in blood pressure occurred. Division of the nerves, on the other hand, without arrest of the circulation in the limb did not prevent the onset of shock.

Others who have repeated Cannon and Bayliss' experiment have not been able to obtain the same results, some indeed have reported findings quite the reverse.

Modern experimental work on shock

Shock has been induced experimentally in several ways, the most common methods employed being mechanical trauma, e.g., pounding the limb of an anesthetized animal with a mallet, or bleeding an animal to the extent of from 40 to 60 per cent of its calculated blood volume and reducing its blood pressure thereby to 30 mm. Hg or less.

Blalock caused shock in dogs by inflicting severe contused injuries to a limb by beating with a mallet. From his results he concluded that a circulating toxic substance is not responsible for the development of the shock syndrome, that transudation of plasma from vessels remote from the traumatized tissues does not occur, and that the reduction of blood volume in traumatic shock can be entirely accounted for by the escape of blood and plasma from the vessels at the *site of the injury*. He bases this conclusion upon the finding that the quantity of blood fluid which accumulates within the part, as estimated by comparing its weight with that of the sound limb, is sufficient to cause the fall in blood pressure. In some experiments as much as a third or a half of the total blood volume was calculated to have entered the traumatized area. The loss of this amount of circulating fluid in itself is sufficient to cause death. From their results Blalock and his followers conclude that there is no essential distinction to be drawn between shock and the state induced by hemorrhage. This has come to be known as the *theory of local fluid loss*.

An animal in shock due to severe hemorrhage can usually be saved by immediate or early transfusion of blood, or even of a good blood substitute, such as gelatin, isinglass or dextran. The blood pressure is quickly raised from the shock level to normal. This period following hemorrhage in which the animal's condition is amenable to treatment is called the "reversible stage" of hemorrhagic shock. At this time the minute vessels are hyperactive to

local stimulation, as first pointed out by Zweifach, Shorr and associates, who have also demonstrated that a pressor principle (vaso-excitor material, or briefly VEM) derived from the kidney appears in the circulation together with a vasodepressor material (VDM) of hepatic origin. If left untreated, the shocked animal passes into a state in which treatment is without avail. This is called the "irreversible stage." VDM is present alone or in excess in the circulation, and the reactivity of the vessels is subnormal. VEM and VDM are produced under anoxic conditions. In the reversible stage of shock, VDM is destroyed by an enzyme system in the liver. In the irreversible stage this mechanism fails to function and VDM accumulates in the blood.

According to Zweifach and associates, shock caused in other ways, e.g., muscle trauma, also exhibits reversible and irreversible stages, and the appearance of VEM and VDM in the blood stream. Usually, however, depression of the blood pressure to the shock level by severe mechanical injury is not amenable to treatment by transfusion, the blood pressure is raised only temporarily, in marked contrast to the response in early hemorrhagic shock (Taylor and Moorehouse, Taylor and Drummond). That is, shock due to severe trauma may be irreversible from the start. This suggests that a factor other than, but in addition to, local fluid loss is responsible for the fall in blood pressure in shock. This factor may well be the vasodepressor material just mentioned (see Markowitz).

In only a very few instances of traumatic shock in man has deficiency of adrenal cortical function been found to play a part. But since in almost any kind of stress, cortical principles are drawn upon in increasing amounts, any pre-existing insufficiency of the adrenal cortex will lead to early exhaustion of its hormones and constitute an aggravating element. The manner in which a greater demand is made for corticoids in shock and other forms of stress is unknown.

A consideration of possible nervous factors in the development of traumatic shock

There is evidence that nervous influences—afferent impulses from the traumatized tissues—may play a part in the development of shock. Cannon suggested that in certain instances the primary cause is overstimulation of the sympathoadrenal system, and Erlanger and Gasser had shown that shock could be produced by the prolonged administration of adrenaline. Freeman, working in Cannon's laboratory, found that the continuous infusion of adrenaline in physiological dosage

(0.001 to 0.006 mg per kilogram of body weight per minute) over a period of two hours caused a fall in plasma volume of from 10 to 29.5 per cent. A reduction in blood pressure to considerably below the normal level occurred upon stopping the injection. Adrenaline, as we know, causes vasoconstriction of the splanchnic and cutaneous areas and dilatation of the vessels of the skeletal muscles (ch. 59). The prolonged arteriolar constriction in the cutaneous and splanchnic areas, it is conceived, by inducing asphyxia of the capillary walls renders them more permeable. The vasodilatation in the muscles by increasing the capillary filtration pressure will increase transudation of fluid in these situations. Other conditions, such as "sham rage," which excite the sympatho-adrenal mechanism were also found to reduce the blood volume (by from 11 to 35 per cent) and cause a fall in blood pressure (to 60 mm Hg). Sham rage induced after the removal of the sympathetic chains does not cause either a reduction in blood volume or a fall in blood pressure.

The researches of O'Shaughnessy and Slome point in the same direction. These workers found that blocking the nervous pathways by spinal anesthesia delayed or prevented the onset of shock in animals whose limbs had been traumatized.

The treatment of traumatic shock resolves itself into a correction of the abnormal physiology indicated in the foregoing discussion, namely, the restoration of the circulating blood volume by blood transfusion, the administration of fluids and the protection of the patient from cold, pain, anxiety and apprehension. Overheating the patient by means of hot water bottles or radiant heat is to be avoided since it induces dilatation of superficial vessels and thus increases the capacity of the vascular bed. The use of agents, e.g., pituitrin, which constrict the capillary vessels, is indicated. Adrenaline is of little or no value, its effect is evanescent and is exerted mainly on the arterioles whose tone in shock is already high. If adrenal cortical function is subnormal the administration of cortisone may be beneficial.

Burn shock. An extensive burn may cause death from primary shock which follows almost immediately after the injury has been received, from secondary shock which supervenes in from 2 to 6 hours after the burning, or from toxemia which develops in from 48 to 72 hours. We are concerned here only with the cause of secondary shock due to burns.

An outstanding feature of burn shock wherein it differs from shock caused by mechanical injury

extreme degree of hemoconcentration which as a result of plasma loss from the burned area and into the tissues surrounding the burn. In an extensive burn, according to Underhill, 10 per cent of the total plasma volume may be removed from the circulation within a few hours. The state following burns is attributed by many observers largely to the reduction in blood volume and to the increased viscosity of the blood. It is estimated the fluid loss by burning one animal while it was on a weighing device tipped towards the burned side as fluid was exhaled. The average amount of fluid lost measured in this way was 2.2 per cent of the weight. The increase in weight of the burned animal commenced almost immediately following the burn and continued at a rapid rate for some time. The blood pressure, however, remained near normal level until a short time before death, at which time it collapsed rather suddenly.

The question of a *toxic factor* in burn shock is raised for the reason that a distinction has not yet been drawn between the stage of shock and the later stage of toxemia which results from absorption of proteolytic products derived from the burnt tissues or of bacterial origin. Robertson and Boyd were among the first to investigate the question of a burn toxin. They found that an alcoholic extract of burned skin, but not of normal skin contained a toxic agent which caused death when injected into normal animals. The blood or red cells, but not the plasma, of burned animals also contained the toxic agent. They concluded that a toxin formed from the burned tissues was absorbed into the blood stream and carried by the red cells. Vogt also found that if a burned area of skin from one animal (guinea pig) were excised and transplanted to another, the first animal survived but the second died. This experiment, however, as well as the work of Robertson and Boyd, is pertinent to the question of burn toxemia which occurs at a later stage than to the development of shock itself. The finding of a high concentration of histamine in the blood of burned patients has raised the question of a toxic factor in burn shock. Barsoum and Gaddum determined the histamine concentration of the blood in a number of patients suffering from extensive burns and found it several times higher than the normal level. However, no clear correlation in time between histamine concentration in the blood and the onset of shock was found. Rose and Brown ob-

served that the course of the blood histamine concentration following burns could be divided into three stages, (1) an early increase, this was not invariable, (2) a marked decrease during the period of edema and plasma loss, i.e., *during the stage of secondary shock*, and (3) a return to normal or above normal as the edema subsides and the patient is improving. It would not appear from these results that histamine is the responsible agent in the development of shock following burns.

Certain changes found at autopsy in burned subjects, e.g., damage to liver and kidney and intestinal ulceration have suggested a toxic factor to many, but such findings are probably associated with the toxemic stage rather than with the stage of shock.

The most effective measure for combatting burn shock is the transfusion of large volumes of plasma or serum (up to 5000 cc. in 24 hours) or a suitable substitute, with the object of restoring the blood volume and coincidentally of reducing the hemoconcentration. Benefit has been reported by some observers (e.g., Wilson, Rhoades, Lee and their colleagues) from the use of adrenal cortical extracts. The cortical hormone is said to decrease capillary permeability and thereby to reduce the plasma loss.

The "crush syndrome" A person who has had a limb held and compressed for some hours by a heavy object such as a beam or a pile of rubble may pass into a state resembling shock some time after he has been released and removed to hospital. A number of such cases have been reported in England following air raids. But a closer study has revealed that they do not present the typical features of wound shock, death has been due to uremia. The part after its release may not appear to be severely damaged and the patients seemed at first to do well after reaching hospital. Marked edema of the limb developed later, accompanied by oliguria. The urine is brownish in color and contains dark granular casts, it gives the benzidine reaction due to the presence of myoglobin (p. 58) derived from the damaged muscles. Complete suppression of urine ultimately supervenes. The condition has been studied by Bywaters and his associates, who attribute the renal effects to damage of the tubules (nephrosis) by myohematin, or a substance closely associated with it and derived from the injured muscles. The anuria would, therefore, be caused in a manner analogous to that resulting from the transfusion of incompatible blood (p. 44), the difference being that in one the pigment is liberated from muscle, in the other from hemolysed erythrocytes.

local stimulation, as first pointed out by Zweifach, Short and associates, who have also demonstrated that a pressor principle (vaso-excitor material, or briefly VEM) derived from the kidney appears in the circulation together with a vasodepressor material (VDM) of hepatic origin. If left untreated, the shocked animal passes into a state in which treatment is without avail. This is called the "irreversible stage." VDM is present alone or in excess in the circulation, and the reactivity of the vessels is subnormal. VEM and VDM are produced under anoxic conditions. In the reversible stage of shock, VDM is destroyed by an enzyme system in the liver. In the irreversible stage this mechanism fails to function and VDM accumulates in the blood.

According to Zwenach and associates, shock caused in other ways, e.g., muscle trauma, also exhibits reversible and irreversible stages, and the appearance of VEM and VDM in the blood stream. Usually, however, depression of the blood pressure to the shock level by severe mechanical injury is not amenable to treatment by transfusion, the blood pressure is raised only temporarily, in marked contrast to the response in early hemorrhagic shock (Taylor and Moorehouse, Taylor and Drummond). That is, shock due to severe trauma may be irreversible from the start. This suggests that a factor other than, but in addition to, local fluid loss is responsible for the fall in blood pressure in shock. This factor may well be the vasodepressor material just mentioned (see Markowitz).

In only a very few instances of traumatic shock in man has deficiency of adrenal cortical function been found to play a part. But since in almost any kind of stress, cortical principles are drawn upon in increasing amounts, any pre-existing insufficiency of the adrenal cortex will lead to early exhaustion of its hormones and constitute an aggravating element. The manner in which a greater demand is made for corticoids in shock and other forms of stress is unknown.

A consideration of possible nervous factors in the development of traumatic shock

There is evidence that nervous influences—afferent impulses from the traumatized tissues—may play a part in the development of shock. Cannon suggested that in certain instances the primary cause is overstimulation of the sympathoadrenal system, and Erlanger and Gasser had shown that shock could be produced by the prolonged administration of adrenaline. Freeman, working in Cannon's laboratory, found that the continuous infusion of adrenaline in physiological dosage

(0.001 to 0.006 mg. per kilogram of body weight per minute) over a period of two hours caused a fall in plasma volume of from 10 to 29.5 per cent. A reduction in blood pressure to considerably below the normal level occurred upon stopping the injection. Adrenaline, as we know, causes vasoconstriction of the splanchnic and cutaneous areas and dilatation of the vessels of the skeletal muscles (ch. 59). The prolonged arteriolar constriction in the cutaneous and splanchnic areas, it is conceived, by inducing asphyxia of the capillary walls renders them more permeable. The vasodilatation in the muscles by increasing the capillary filtration pressure will increase transudation of fluid in these situations. Other conditions, such as "sham rage," which excite the sympatho-adrenal mechanism were also found to reduce the blood volume (by from 11 to 35 per cent) and cause a fall in blood pressure (to 60 mm. Hg). Sham rage induced after the removal of the sympathetic chains does not cause either a reduction in blood volume or a fall in blood pressure.

The researches of O'Shaughnessy and Slome point in the same direction. These workers found that blocking the nervous pathways by spinal anesthesia delayed or prevented the onset of shock in animals whose limbs had been traumatized.

The treatment of traumatic shock resolves itself into a correction of the abnormal physiology indicated in the foregoing discussion, namely, the restoration of the circulating blood volume by blood transfusion, the administration of fluids and the protection of the patient from cold, pain, anxiety and apprehension. Overheating the patient by means of hot water bottles or radiant heat is to be avoided since it induces dilatation of superficial vessels and thus increases the capacity of the vascular bed. The use of agents, e.g., pituitrin, which constrict the capillary vessels, is indicated. Adrenaline is of little or no value, its effect is evanescent and is exerted mainly on the arterioles whose tone in shock is already high. If adrenal cortical function is subnormal the administration of cortisone may be beneficial.

Burn shock. An extensive burn may cause death from primary shock which follows almost immediately after the injury has been received, from secondary shock which supervenes in from 2 to 6 hours after the burning, or from toxemia which develops in from 48 to 72 hours. We are concerned here only with the cause of secondary shock due to burns.

An outstanding feature of burn shock wherein it differs from shock caused by mechanical injury

is the extreme degree of hemoconcentration which occurs as a result of plasma loss from the burned surface and into the tissues surrounding the burn. In an extensive burn, according to Underhill, 70 per cent of the total plasma volume may be lost from the circulation within a few hours. The shock state following burns is attributed by many observers largely to the reduction in blood volume and to the increased viscosity of the blood. Harkins estimated the fluid loss by burning one side of an animal while it was on a weighing device which tipped towards the burned side as fluid accumulated. The average amount of fluid lost as measured in this way was 22 per cent of the body weight. The increase in weight of the burned side commenced almost immediately following the burning and continued at a rapid rate for some time. The blood pressure, however, remained near the normal level until a short time before death, when it collapsed rather suddenly.

The question of a *toxic factor* in burn shock is confused for the reason that a distinction has not always been drawn between the stage of shock and the later stage of toxemia which results from the absorption of proteolytic products derived from the burnt tissues or of bacterial origin. Robertson and Boyd were among the first to investigate the question of a burn toxin. They reported that an alcoholic extract of burned skin but not of normal skin contained a toxic agent which caused death when injected into normal animals. The blood or red cells, but not the plasma, of burned animals also contained the toxic agent. They concluded that a toxin formed in the burned tissues was absorbed into the blood stream and carried by the red cells. Vogt also reported that if a burned area of skin from one animal (guinea pig) were excised and transplanted to another, the first animal survived but the second died. This experiment, however, as well as those of Robertson and Boyd, is pertinent rather to the question of burn toxemia which occurs at a later stage than to the development of shock itself. The finding of a high concentration of histamine in the blood of burned patients has revived the question of a toxic factor in burn shock. Barsoum and Gaddum determined the histamine concentration of the blood in a number of patients suffering from extensive burns and found it several times higher than the normal. However, no clear correlation in time between the histamine concentration in the blood and the onset of shock was found. Rose and Brown ob-

served that the course of the blood histamine concentration following burns could be divided into three stages, (1) an early increase, this was not invariable, (2) a marked decrease during the period of edema and plasma loss, i.e., *during the stage of secondary shock*, and (3) a return to normal or above normal as the edema subsides and the patient is improving. It would not appear from these results that histamine is the responsible agent in the development of shock following burns.

Certain changes found at autopsy in burned subjects, e.g., damage to liver and kidney and intestinal ulceration have suggested a toxic factor to many, but such findings are probably associated with the toxemic stage rather than with the stage of shock.

The most effective measure for combatting burn shock is the transfusion of large volumes of plasma or serum (up to 5000 cc in 24 hours) or a suitable substitute, with the object of restoring the blood volume and coincidentally of reducing the hemoconcentration. Benefit has been reported by some observers (e.g., Wilson, Rhoades, Lee and their colleagues) from the use of adrenal cortical extracts. The cortical hormone is said to decrease capillary permeability and thereby to reduce the plasma loss.

The "crush syndrome" A person who has had a limb held and compressed for some hours by a heavy object such as a beam or a pile of rubble may pass into a state resembling shock some time after he has been released and removed to hospital. A number of such cases have been reported in England following air raids. But a closer study has revealed that they do not present the typical features of wound shock, death has been due to uremia. The part after its release may not appear to be severely damaged and the patients seemed at first to do well after reaching hospital. Marked edema of the limb developed later, accompanied by oliguria. The urine is brownish in color and contains dark granular casts, it gives the benzidine reaction due to the presence of myoglobin (p. 58) derived from the damaged muscles. Complete suppression of urine ultimately supervenes. The condition has been studied by Bywaters and his associates, who attribute the renal effects to damage of the tubules (nephrosis) by myohematin, or a substance closely associated with it and derived from the injured muscles. The anuria would, therefore, be caused in a manner analogous to that resulting from the transfusion of incompatible blood (p. 44), the difference being that in one the pigment is liberated from muscle, in the other from hemolysed erythrocytes.

An important factor in the production of the "crush syndrome" is probably ischemia of the renal cortex, which seems to render the tubules more susceptible to the action of the blood pigment. Trueta and his associates found that a diversion of blood from the greater part of the renal cortex to the deeper vessels, accompanied by narrowing of the renal artery, occurs in tourniquet shock. Prolonged stimulation of the central end of an afferent nerve (sciatic) causes a similar "by passing" of the cortical glomeruli (p. 478).

The nearest approach to the "crush syndrome" which can be made in experimental animals is by the production of *tourniquet shock*. When both hind limbs of an animal such as the dog are made ischemic by compression for three or four hours, and the limbs then released, the members become swollen and a steady fall in blood pressure follows. The shock cannot be explained simply upon the basis of the loss of blood fluid into the tissues of the swollen part. Death occurs within a few hours. There is a rise in the non protein nitrogen of the blood and there may be the passage of blood stained urine before death. In experiments upon cats Eggleton found that an extract prepared from the affected muscles caused impairment of renal function (reduction of creatinine clearance) when injected into normal animals. Cortical ischemia may occur in other forms of injury, it has been observed in man in severe hemorrhage, though the renal medulla was still supplied with blood.

Blast injury. The explosive blast of a bomb or shell may cause death without any sign of external injury. But hemorrhages into the lungs, beneath the pleura or into the tissues of the liver and other abdominal viscera may be found at autopsy. Actual laceration of pulmonary tissue or of the abdominal organs may occur. The effects of blast have been studied experimentally by Barcroft and by Zuckerman. Barcroft attributes the pulmonary trauma to the sudden distension of the lungs and the rupture of the lung tissue by the blast of air caused by the explosion. Zuckerman believes it to be due to the impact of the percussion wave set up by the explosion upon the body surface. The first theory does not explain the abdominal lesions that are so frequently observed. Nor is the second theory able to explain satisfactorily the pulmonary lesions, for one would expect the thoracic walls to protect the lungs from the force of the blast. It is probable that, in most instances, both factors play a part in blast injuries.

Anaphylaxis, anaphylactic shock

The anaphylactic reaction is one which follows the administration of a foreign substance (most frequently protein in nature) to an animal which has been sensitized to it by a previous dose. It is one of the manifestations of an antigen antibody response. In the dog and guinea pig, an interval of from 12 to 14 days

must elapse between the two doses, in order that the maximum effects of the second dose may be produced.

The anaphylactic reaction, which in its severer manifestations is called anaphylactic shock, shows pronounced species differences. *In the dog* it is marked by dyspnea, vomiting, diarrhea, salivation, a profound fall in blood pressure ending in circulatory failure and death. The hepatic veins are strongly constricted, the liver and intestinal vessels are engorged and there is an increased flow of lymph from the thoracic duct. The blood is incoagulable as a result of the liberation of heparin from the liver (Waters and associates). The histamine content of the liver of sensitized animals is increased, but is reduced below the normal in anaphylaxis. Thus, the major symptoms of anaphylactic shock in the dog are centered in the liver. Nevertheless, many of the features of anaphylactic shock can be produced in the hepatectomized dog. *In the guinea pig*, the anaphylactic reaction consists of a powerful contraction of the muscles of the bronchioles. There is extreme dyspnea, death is due to asphyxia. The effect can be demonstrated in the perfused isolated lungs of a sensitized animal by the addition of the antigen to the perfusion fluid. The histamine content of the blood and lungs is increased to several times the normal.

In the rabbit, the reaction may be general or purely local. When the foreign substance is injected subcutaneously into a sensitized animal, the skin and subcutaneous tissue at the site of the injection become edematous and swollen, a sterile abscess or slough appears. This was originally described by Arthus and is known as the Arthus phenomenon. When the antigen is administered intravenously, the blood pressure falls abruptly, the respirations become rapid, but there is no dyspnea. The bladder and intestine are evacuated. The animal may die within a few minutes from dilatation and failure of the right ventricle. The failure of the heart is secondary to the increased resistance in the pulmonary circuit caused by constriction of the arterioles. The arterioles in other parts of the vascular system are also constricted and emboli composed of clumps of leucocytes may be seen blocking the pulmonary and systemic capillaries. This, no doubt, accounts for the leukopenia which is found in the peripheral blood stream.

The evidence indicates, with little doubt, that, in anaphylactic shock, the antibody antigen reaction takes place in or on the tissue cells and not in the blood plasma. Dale showed for example, that when the uterus was removed from a sensitized guinea pig and its vessels freed from all traces of blood, it gave the typical anaphylactic contraction when the antigen was added to the bath in which it was immersed. Manwaring also found that the blood of the sensitized animal could be replaced by blood from a normal animal without affecting the first animal's sensitivity. Sensitivity can be passively transferred by the injection of the serum of a sensitized animal into a normal animal. A latent period elapses

between the injection and the development of the passive sensitization which is presumably required for the fixation to the tissue cells of the antibody transferred in the injected blood. The interval of from 12 to 14 days, as mentioned above, which must elapse in order for the maximum effects of the antigen to become manifest is used, presumably, for the production of the antibody and its attachment to the tissue cells.

It is now generally accepted that the antigen-antibody reaction in some way brings about the liberation of histamine from the affected tissues and that the action of this amine is responsible for the anaphylactic manifestations. Nearly all the features, as seen in these three species, can be explained upon such a basis. In the dog and guinea pig, anaphylactic shock is associated with a rise in the histamine concentration of the blood and, although the whole blood of the rabbit shows no increase and is often reduced, the amine passes from the white cells (which contain it in especially large amounts) into the plasma. The species peculiarities of the anaphylactic manifestations can be accounted for largely by the amount of smooth muscle in the reactive tissues of these three species and by its susceptibility to the action of histamine. In the dog, the smooth muscle of the hepatic veins is especially well developed. In the guinea pig, the bronchioles are particularly susceptible to stimulation by histamine and, in the rabbit, the pulmonary arterioles show unusually thick muscular coats.

The mechanism of histamine liberation from the sensitized cells is unknown, but certain experimental results suggest a proteolytic enzyme action. It has been found, for example, that trypsin, which exerts an action on smooth muscle resembling that of histamine or anaphylactic shock, causes the release of histamine from the perfused lungs of guinea pigs.

Thus, the liberation of histamine from the cells of the sensitized animal might be due to the action of the antigen in lowering antitryptic activity and thereby permitting intracellular trypsin to exert its proteolytic action. Peptones, whose manifestations are closely similar to those of anaphylactic shock (see below) would also be formed.

Though histamine liberation appears undoubtedly to be the major factor in the production of the phenomena of anaphylaxis, certain observations which do not conform to the histamine theory suggest that some other factor or factors are involved. Minute doses of histamine, for example, cause contraction of the isolated rat's uterus, whereas large doses of antigen are required to produce even a weak contraction. Also, the isolated guinea pig's uterus, poisoned by high concentrations of histamine, responds to a further dose by relaxation, but by contraction to a further dose of antigen.

Peptone solution, injected intravenously, produces in the dog effects almost identical with those of anaphylactic shock, including incoagulability of the blood due to the liberation of heparin from the liver. Sensitization by a previous dose is not necessary, however. Adding peptone to rabbit blood *in vitro* causes the liberation of histamine from the cells into the plasma.

In man, fatal anaphylactic shock may follow the injection of horse serum into a person who has been sensitized by a previous administration. Allergic reactions, in general, show many similarities to anaphylactic reactions and many observations suggest that the two are fundamentally allied, although in the case of allergic reactions, sensitization by an earlier exposure cannot always be demonstrated, the reaction appearing upon the first known contact with the foreign substance.

CHAPTER 28

SPECIAL FEATURES OF THE CIRCULATION IN DIFFERENT REGIONS

THE CAPILLARY CIRCULATION

The pattern of the vascular bed lying between the arterioles and the venules

Chambers, Zweifach and their associates have studied the architecture of the capillary system of vessels extensively and with great ingenuity in the mesentery and omentum of the dog and in the mesoappendix of the rat. Contrary to previous ideas that all vascular channels lying between the arteriole and the larger venules were formed of simple endothelial tubes, these observers find that smooth muscle fibers extend for some distance beyond the terminal arteriole along a vessel centrally placed in respect to neighboring capillaries. This vessel, which runs from arteriole to venule, is called the *metarteriole*. From it the capillaries proper or *true* capillaries—simple endothelial

tubes—are given off, but after a short course, loop back and re-enter the central channel nearer to its venous end (see fig 28 1). The capillary at the junction of its efferent limb with the metarteriole is provided with strands of smooth muscle. This *precapillary* sphincter, as it is called, is capable, by widening or narrowing the outlet from the central channel, of controlling the flow through the capillary. No sphincter guards the junction of the afferent limb of the capillary with the metarteriole, nor does the wall of the more distal (venous) portion of the latter possess muscular elements. Connecting vessels running from arteriole to venule, arteriovenous anastomoses, are also frequently to be seen. Through these channels blood can be shunted upon occasion and a capillary area short-circuited.

The metarterioles show rhythmical variations in caliber and, consequently, in the blood flow to the capillaries beyond. When the tissue is in a resting state, the constrictor phase of this rhythm predominates and the precapillary sphincters may be completely closed. When the tissue becomes active, the dilator phase of the metarteriole predominates and the precapillary sphincters are open. The metarteriolar wall and the sphincters are highly sensitive to local changes, any mechanical irritation causing dilatation, and to agents such as adrenaline and histamine. The former causes constriction of the sphincters in a dilution of 1:10 million and of the metarterioles in a dilution of 1:4 million. Histamine, in a concentration of from 1:4000 to 1:8000, is without effect upon the arterioles, but dilates the metarterioles and precapillary sphincters.

The observations of several investigators in the past have pointed to the capillaries being capable of regulating their size quite independently of the state of the arterioles which feed them. Stricker (1865), for example, reported active contractions and dilatations of capillaries in the nictitating membrane of the frog. Roy and Brown found that individual, capillaries lying close together and apparently supplied with blood from a common arteriole, required different pressures to obliterate their respective lumina. Steinhilber and Kahn ob-

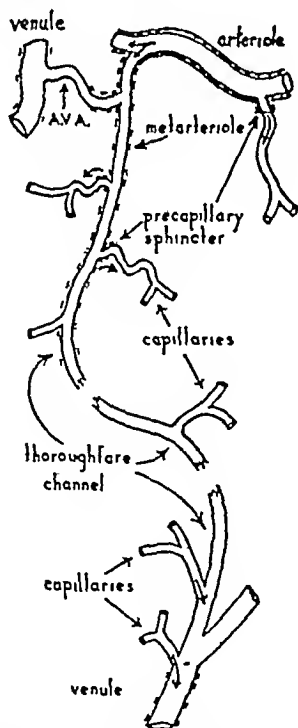


FIG 28 1 Diagram of a functional unit of the capillary bed, AV A, metarteriolar venular anastomosis (Chambers and Zweifach)

tained more direct evidence of independent capillary movements by electrical stimulation of excised tissues (nictitating membrane of the frog and mesentery of the cat). Constriction of individual capillaries was clearly demonstrated. These observers also reported that stimulation of the cervical sympathetic in the living animal caused contraction of the capillary walls in an area *from which the circulation had been occluded*. Scant attention apparently was paid to these observations. Until the more recent work of Lewis and of Krogh and their associates, few physiologists believed that *active* changes in capillary diameters could occur. The capillaries were considered to play a purely *passive* role; changes in their capacity were believed to depend entirely upon the distension or collapse of their elastic walls in strict accordance with the diameters of the arterioles at the moment, and consequently upon the quantity of blood which they received from the arterial side.

If all vessels lying between the arterioles on the one hand and the smallest veins on the other are classed as capillaries, the view that these vessels are merely inert tubes with distensible and elastic walls, but otherwise incapable of altering their calibers, is now generally admitted to be wrong. That changes in the capillary circulation can occur quite independently of the state of the arterioles is now a firmly established fact. However, in the capillary areas studied by Chambers and Zweifach, the so called *true* capillaries are relatively immobile, behaving to a large extent as passive conduits. The earlier evidence for independent movement of the capillaries, as secured by Lewis and his associates and by Krogh, will be briefly reviewed. But it should be remembered that all the minute vessels lying distal to the arterioles were considered to be of capillary structure, that is, composed simply of endothelial tubes, the true capillaries of Chambers and Zweifach. The presence of smooth muscle in the more proximal vessels (the metarterioles between the arterioles and the simple endothelial vessels) was then unsuspected, and it is very probable that the very active movements described by these earlier workers in the tissues which were under examination belonged to vessels containing smooth muscle.

(1) Cotton, Slade and Lewis showed that reddening of the skin of the arm could be induced by mechanical stimulation, after the circulation had been arrested by means of a blood pressure cuff. The dilatation of the

superficial vessels, it was argued, must have been the result of relaxation of the capillary wall and not of passive distension, since after the blood in the part had become stagnant, and venous and arterial pressures equal, arteriolar dilatation would have drained blood from, rather than forced it into the capillaries. A white line was also obtained upon drawing a point lightly over the skin, which would only be satisfactorily explained by capillary constriction. (2) Krogh observed the capillaries in the tongue muscles of the frog beneath the microscope, and showed that they contracted when stimulated mechanically, as by touching them with a stiff hair. Dilatation could be induced toward the venous end of the capillary while the diameter of the arterial end remained unaltered, the dilated vessel filled from the vein. Drawing the hair along the capillary toward the arteriole was followed by a progressive opening up and filling of the former vessel. As the dilatation reached the arteriole the arterial blood was admitted and the flow in the capillary became reversed. (3) When the arteries to the frog's tongue were occluded, active dilatation of the capillaries was induced by mechanical stimulation. If, on the other hand, the capillaries were constricted they remained so, against an arterial pressure of several millimeters of mercury. Lewis also showed that the capillaries of the human skin were capable of remaining contracted against an arterial pressure of approximately 100 mm of mercury. (4) Various chemical substances, e.g., acids (lactic, CO_2), urethane, iodine, cocaine, silver nitrate and chloroform cause dilatation of the capillaries, some of these, e.g., urethane and iodine, have no effect upon the arterioles. Alkalis and oxygen cause capillary constriction. Though adrenaline (in ordinary dosage) causes constriction of both arterioles and capillaries in man, in the frog the arterioles are constricted while the capillaries are often simultaneously dilated. Histamine induces dilatation of both types of vessels in man but, as shown by Dale and Richards, causes arteriolar constriction and capillary dilatation in the cat. Acetylcholine and adrenaline applied directly to the arterioles cause dilatation and constriction, respectively, yet, during these arteriolar changes, scarcely noticeable alterations in capillary diameters occur.

Krogh found that there was a great variation in the number of capillaries which might be pervious at a given moment. The number of patent vessels was directly related to the activity of the tissue at the time, the number of open vessels being much fewer in resting than in active muscle. When, for example, India ink was injected into the vessels and portions of tissue excised before and after excitation, the increase in number of open vessels in the latter instance was very striking. The resting muscle in the intact

animal is pale and only relatively few vessels can be seen. Others may be observed to be contracted to such small diameters that a red cell is unable to enter the narrowed channel without becoming squeezed out of shape. A cell pressed into a elongated or sausage like form could frequently be seen within the vessel or, in some instances, imprisoned by the complete closure of the capillary lumen at each end. In other instances, constriction of the vessel excludes the corpuscles entirely from a neighboring capillary channel, the plasma alone being permitted to enter. This phenomenon has been termed "plasma skimming." Upon stimulating the muscle the picture was entirely changed, capillaries hitherto invisible, since their lumina apparently had been obliterated by tonic contraction, sprang into view, while other vessels which had been partially closed dilated fully. Krogh considers that these changes play an important rôle in the mechanism for the regulation of the oxygen supply to the tissues. During rest a comparatively large area of tissue is fed with blood from a single central capillary. During activity, as more capillaries become patent, the area of tissue dependent upon each one becomes correspondingly reduced (see fig 32.5, p 375). Richards also observed a great increase in the number of capillary tufts in the frog's kidney, when the activity of the organ increased (see p 452). From these observations upon frog's muscle and kidney, conclusions cannot, however, be drawn regarding the blood supply to tissue in general, since Lewis has found that, in the human skin at any rate, the capillary pattern changes very little from time to time, all the vessels are open whether the blood flow is small or large. Lewis concludes that augmentation of the blood flow through the skin is brought about through the greater dilatation of pre-existing channels rather than by the opening of capillaries that had been previously closed.¹

The elements in the walls of the true capillaries responsible for their independent movements

Though the true capillaries, according to the observations of Chambers and Zweifach, are not capable of movements strong enough to materially

¹ Bordley and associates disagree with this conclusion, for they observed intermittency of flow in the capillaries of the skin over the tibia. Even two capillaries arising from a single arteriole were seen to alternate with one another in permitting the passage of corpuscles.

alter the capillary blood-flow, active changes in the calibers of these vessels to a very moderate degree do occur. The endothelial cells of the true capillaries have been seen to change their shape from time to time, but such movements seemed too weak to alter the blood-flow, and were a response to rather than a cause of changes in capillary blood flow. Nor did Clark and Clark, in studies of the capillaries in the rabbit's ear, see any indication that the endothelial cells were capable of any strong contraction. Much discussion has, therefore, been aroused and several suggestions offered with respect to even the moderate active changes in caliber of the true capillaries which has been observed. The movements have been attributed to contraction of the Rouget cells (p 138), to a constricting action of fibers in the pericapillary sheath, to imbibition of fluid and swelling of the endothelial cells or to bulging of the cell nucleus into the vascular lumen without change in outside capillary diameter.

The Rouget cells found in relation to the capillary wall in amphibians are not, as was previously believed, concerned in the reactions. Clark and Clark give evidence for the Rouget cells being simply wandering connective tissue elements which have come to rest upon the capillary wall. They have observed movements of the endothelial wall independently of the Rouget cells, and Florey and Carleton describe cells in mammals which are probably analogous to Rouget cells, but think that they are too few in number to be of any importance in the capillary movements. Moreover, they state that a clear space separates these cells from the capillary wall. The Rouget cells are probably reticulo-endothelial elements.

The constitution of the capillary wall

The capillary wall proper is composed of a single layer of endothelial cells, and is about the most tenuous structure conceivable. It does not exceed half a micron in thickness. The endothelial tube is enveloped by a delicate sheath of fine fibrils, derived from, or closely related to, the surrounding connective tissues. These fibers, which are stained darkly by silver salts, are embedded in a gel-like matrix, and serve to support the outer aspect of the capillary wall. The cement substance binding the endothelial cells together at their edges, and which also stains with silver salts (fig 28.2), is being continually washed away and renewed again (Chambers and Zweifach). When it is blackened by

Snash

the application of silver salt, parts of the blackened material can be seen to become detached and be swept away in the blood current, yet the integrity of the cementing material is apparently maintained, for no change in permeability of the wall is evident. The endothelial cell is believed to be responsible for the production and continuous renewal of the intercellular cement, Chambers and Zweifach suggest that this is an essentially important function of the cell. Calcium is required for the maintenance of the cement substance in a normal state. When the capillaries are perfused with a calcium-free fluid, softening of the cement occurs and the capillary wall becomes excessively permeable. Softening of the cement is indicated by increased "stickiness", which can be demonstrated by adding carbon particles to the perfusion fluid. The carbon grains accumulate and adhere tenaciously to the intercellular lines, but not to the surfaces of the endothelial cells. Lowering the pH of the perfusion fluid was observed to soften and increase the stickiness of the cement in a similar way.

It is believed protein derived from the plasma proteins is adsorbed upon the inner aspect of the capillary wall to form a thin lining. Removal of this protein layer, as by a perfusate of saline or of gum acacia solution, causes edema, due to leakage of fluid through the capillary membrane, but upon the addition of serum to the perfusion fluid, capillary permeability is reduced again to normal. The adsorbed protein is supposed to act by filling the pores of the intercellular cement.

Capillary permeability

All gaseous and fluid exchanges between the blood and the extravascular fluids take place across the capillary wall. The arteries, even as far, apparently, as the commencement of the true capillaries, are impervious. It has been calculated by Krogh that the total filtering surface by the capillary endothelium in the adult human body amounts to around 6300 square meters (68,000 square feet). The summed areas of the capillary walls in all regions could be imagined, therefore, as constituting an endothelial membrane over 12 miles long and a foot wide, yet so thin that, when tightly rolled, it would form a cylinder of about the thickness of a lead-pencil. The state of the capillary wall with respect to permeability is, therefore, of the utmost importance in maintaining the nutrition of the tissues and in maintaining the

normal distribution of fluids within and without the vascular system.

Some of the aspects of capillary permeability have been touched upon in earlier chapters (pp 32 and 36). It has been stated there that a saline fluid containing proteins in low concentration leaves the capillary at its arterial end and is reabsorbed minus the protein (which enters the lymph vessels) at the venous end. Capillary dilatation with increased blood flow through a capillary area raises the intracapillary pressure and increases the filtration rate, while constriction of the capillary bed favors reabsorption. An increase in filtering area may be brought about by dilatation of previously patent capillaries, or by increasing the number of open capillaries. The capillary blood flow and other factors being the same in both instances, the quantity of fluid filtered will be greater in the latter instance.

Considered as a permeable membrane, the capillary wall offers two elements for study—the endothelial cell itself and the cement substance which binds the edges of the cells together. It is to alterations chiefly in the latter substance that changes in permeability are attributed. Though the permeability of the capillary membrane is such that gases, and probably water and electrolytes of small molecular size as well, can diffuse readily through its cells, the transference of protein and other substances of large molecular size takes place entirely, it is believed, through pores in the intercellular cement. The unselective nature of the transference of various materials, such as unchanged plasma protein, argues against the endothelial cell itself being the medium of such transfer. Also, the filtration rate through the capillary membrane is about 100 times the value for the passage of fluid into a living cell, and, according to a calculation by Landis, amounts to 3.7×10^{-2} cc, per cm^2 , per minute at a pressure of 1 atmosphere. It is, therefore, very widely, though

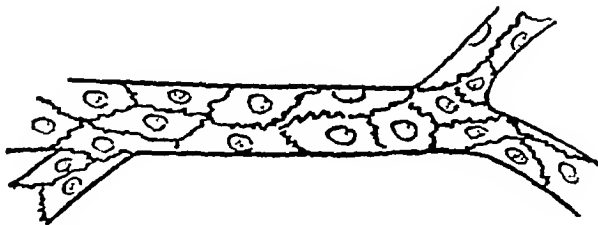


FIG 28 2 Section of capillary wall stained with silver nitrate to demonstrate the intercellular cement. Nuclei of the cells stained with hematoxylin.

perhaps not universally, accepted that the permeability of the capillary wall depends primarily upon the intercellular cement substance. The character of this material is considered in the next section. Capillary permeability is altered by a number of agents. Adrenal cortical steroids, e.g., desoxycorticosterone and corticosterone, calcium salts and a rise in pH reduce permeability, whereas leukotaxin, histamine, tissue extracts, calcium lack and low pH increase it. Capillary permeability is also increased by heat (even of a degree which does not damage the vessel wall) and is diminished by cooling to a temperature which causes no tissue injury. Permeability is of course increased by conditions, e.g., anoxia, vitamin C lack, which result in capillary damage.

The question of a nervous (sympathetic) influence upon capillary permeability is debatable, the balance of evidence appears to be against a nervous regulation.

Hyaluronic acid and hyaluronidase

Hyaluronic acid is a mucopolysaccharide, which possesses the property of binding water and forming a gel like material. It has a very wide distribution in the body, and serves as a cement substance to hold cells together and to bind water in the interstices of the tissues. It also serves as a lubricant in joints and bursae, and in the formation of cushions or buffers. It is present in the *vitreous body* (from which it was first isolated), in *Wharton's jelly* of the umbilical cord, in *synovial fluid*, the *nucleus pulposus* of the intervertebral disc, *skin*, and in the *capsules of hemolytic streptococci* (groups A and B). It also holds together the cells forming the *cumulus* (*cumulus oöphorus*) of the mammalian ovum.

An enzyme which hydrolyzes hyaluronic acid, and known as *hyaluronidase*, is found in a variety of biological fluids and tissues. The action of this enzyme was first demonstrated in subcutaneous tissues by Duran-Reynals, who found that an injection of an extract of testes (bull's) increased the potency of vaccine virus administered subcutaneously. The principle causing this effect was called the "*spreading factor*". Its action was later shown to be due to the breakdown of hyaluronic acid by the enzyme hyaluronidase. The increased effectiveness of the vaccine is, therefore, due to the dissolution of a barrier in the tissues which enables infection to spread more readily.

Hyaluronidase is present in the sperm of most mammals and of many submammalian forms, and

in appropriate concentration is capable of dissolving the cumulus oöphorus surrounding the ovum. It has therefore been thought that its function in the sperm was to disperse this covering of the ovum and therefore to be essential for the penetration of the ovum by the sperm, see chapter 61 where this question is discussed. The enzyme is also found in the venoms of bees and snakes, associated with the toxins of invasive bacteria, e.g., *staphylococcus aureus*, the pneumococcus, hemolytic streptococci and the Welch bacillus, and in the salivary secretions of the leech. Sulfonated and phosphorylated hesperidins were found by Beiler and Markin to inhibit the action of this enzyme, it is also powerfully inhibited by trypsinic acid (rehibin).

Hyaluronidase reduces the viscosity of synovial fluid from its normally high value to that of water. When used in conjunction with subcutaneous injections, it facilitates the spread of the fluid in the tissues by resolving the jelly-like matrix in the interstitial spaces. The bleb or swelling formed by the injected fluid is accompanied by less pain and discomfort, and disappears much sooner, an action which has suggested its use to facilitate the administration subcutaneously of large volumes of fluid and thus to hasten their absorption. It has also been employed clinically in the treatment of sterility resulting from its deficiency in the spermatogenic fluid. Use has also been made of the enzyme to hasten the absorption of drugs and to facilitate local anesthesia.

It has been suggested that the capillary intercellular cement substance is hyaluronic acid, but the observation of Chambers and Zweifach that hyaluronidase, though it attacks the matrix of the pericapillary sheath, and thus weakens the capillary wall, does not alter the permeability of the capillary wall itself, argues against this idea. No softening of the cement, as indicated by "stickiness", was observed, even though the enzyme was applied directly to the capillary wall. Small petechial hemorrhages were seen due to rupture of the capillary deprived of the outer support normally furnished by the pericapillary sheath. Vitamin C lack also results in dissolution of the matrix of the pericapillary sheath (chap 54).

THE INNERVATION OF THE CAPILLARIES

Efferent Fibers Non-medullated nerve filaments were observed by Woollard accompanying the capillaries. They form a loose investment

of the vessels, but none was seen to end actually upon the capillary wall. These nerves are derived from the sympathetic and are extensions from the periarteriolar plexus (p 273), they degenerate after removal of the sympathetic chain. Physiological evidence for constrictor fibers to the capillaries has been obtained by several workers. The observations of Steinach and Kahn have been mentioned (p 309). Hooker observed contraction of the capillaries of the ear upon stimulation of the cervical sympathetic in the cat, which he concluded was a direct effect upon the vessels. Marvin and Harris obtained similar results, but excluded more certainly any possibility of a passive effect upon the capillaries as a result of arteriolar constriction by arresting the circulation to the rabbit's ear before applying the stimulus. Krogh and his colleagues observed contraction of the capillaries of the frog's web when sympathetic fibers were stimulated. Excision of the abdominal sympathetic ganglia, on the other hand, or section of the sciatic, by removing vasoconstrictor tone, caused the capillaries of the web to dilate.

Dilatation of muscle capillaries is probably mainly dependent upon the production of acid metabolites, though it is possible that they also, like the arterioles, are dilated through the medium of sympathetic fibers (p 275).

Engel postulates a permeability-controlling function for the sympathetic based upon experiments in which the vessels of the knee joint were perfused with a dye-containing fluid. Sympathetic stimulation was found to increase, sympathectomy to decrease the excretion of dye into the joint cavity.

AFFERENT FIBERS Woollard observed *medullated* fibers reaching as far as the arterioles but no farther. They were seen to end by collateral branches in the arteriolar wall on the one hand, and in the adjacent subcutaneous tissue on the other. Some of the latter were seen to terminate in Pacinian corpuscles (ch 63). These medullated fibers are sensory and most likely provide the basis for axon reflexes, they probably also convey antidromic dilator impulses from the central nervous system. Doi, for example, caused full dilatation of the arterioles in the frog by means of acetylcholine, upon then stimulating the posterior nerve roots dilatation of the capillaries was observed. Since the blood flow through the arterioles was already maximal the dilatation could

not have been simply a passive effect. This observation has been confirmed by Krogh and his associates who also obtained dilatation of capillaries in the frog's tongue upon mechanical stimulation of the peripheral end of the glosso-pharyngeal. The failure, however, to demonstrate sensory fibers ending upon the capillary wall lends support to the view of Lewis that the dilatation of the cutaneous capillaries is brought about through the intermediary of a chemical substance (H-substance, p 316) liberated at the nerve ending. Stohr also states that a nerve fiber can be traced to the vicinity of only one out of every hundred capillaries, yet all dilate as a result of nerve stimulation. This, of course, is an indication that vessels more remote from the nerve terminal are acted upon by a diffusible chemical substance liberated by the latter.

THE VASCULAR REACTIONS OF THE HUMAN SKIN

The architecture of the vessels of the skin has the general pattern already described on page 308. The arterioles upon approaching the bases of the papillae (i.e., the layer of the corium immediately underlying the epidermis) turn horizontally, and give rise to metarterioles from which originate, in turn, hairpin-shaped endothelial tubes—the *capillary loops*. The proximal or arterial limb of the capillary loop ascends in the papilla and then turns upon itself to form the venous limb. The latter on reaching the base of the papilla joins with the venous limbs of neighboring loops to form a *collecting venule*. The collecting venules anastomose with one another to form a rich plexus—the *subpapillary venous plexus*—which runs horizontally beneath the bases of the papilla. It drains into deeper veins.

The capillary loops, as Lombard first demonstrated, can be seen readily in the living skin under the low power of the microscope. A drop of cedar or paraffin oil is placed upon the skin at the base of the finger nail or other region where the epidermis is thin. The area under observation is illuminated by a powerful beam of light. If the horny layer be removed by blistering, the vessels are rendered more distinct. Under favorable circumstances the vessels of the subpapillary venous plexus and even deeper vessels may also come into view. The vessels at the base of the human finger nail are shown diagrammatically in figure 28.3.

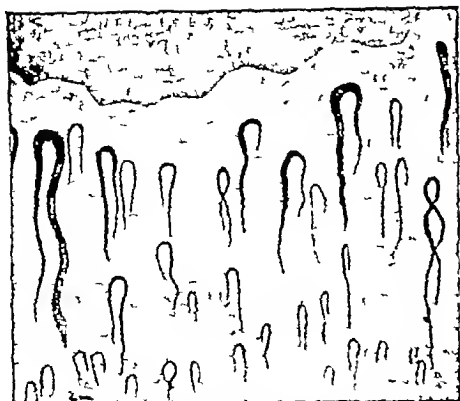


FIG. 28.3 The bed of the finger nail in a healthy subject, showing the capillary loops and the summits of the skin papillae (After Lewis)

The capillary circulation in relation to the color and temperature of the skin

The color of the skin may afford important diagnostic information, and the factors underlying the variations in the tint and depth of color are of considerable interest from a purely physiological point of view. Lewis has made an extended study of the capillaries of the human skin and their reactions to various types of stimulus. The color of the skin is not dependent normally upon the most superficial vessels, i.e., the capillary loops, but upon the subpapillary venous plexus. The vessels of the plexus, though more deeply placed, present a greater area parallel to the skin, whereas the capillary loops are disposed chiefly at right angles to the skin surface. A simple experiment devised by Lewis illustrates the predominant rôle played by the subpapillary venous plexus in determining the skin color. When the superficial vessels were observed beneath the microscope through a glass slide, it was found that as the slide was gradually pressed down upon the skin the venous plexus collapsed first, not until further pressure was applied were the capillary loops obliterated. Blanching of the skin occurred, nevertheless, at the moment when the blood was pressed from the plexus and while the capillary loops could still be seen, subsequent occlusion of the capillary loops did not increase the pallor appreciably.

When the skin is unusually pale and little blood is contained in the superficial vessels (subpapillary venous plexus), the skin is more transparent, and the deeper venous plexuses then contribute largely

to the color of the skin, often adding a leaden tint to the pallor. When the overlying vessels are open and the skin is well supplied with blood these deeper vessels are hidden from view.

Apart from pigmentary effects and assuming the general arterial blood to be normal, the color of the skin, i.e., the dominance of the reddish or of the bluish hue depends upon the extent to which the oxyhemoglobin becomes reduced during the passage of the blood through the cutaneous vessels. The degree of reduction will depend entirely, as a rule, upon the rate of blood flow. When the flow is rapid or slow the blood is, respectively, more arterial or more venous in character. The tint of the skin varies accordingly.

The depth of the skin color, i.e., the intensity of color apart from hue is dependent upon the diameters and the degree of engorgement with blood of the superficial vessels.

So, taking into account both the hue and the depth of color, an intense scarlet color of the skin indicates a normal or increased blood flow and dilated vessels, a deep blue color (see cyanosis, p. 435) accompanies a slowed blood flow and dilated vessels (such as would result from obstruction to the venous trunks). Pallor or a light pink color of the skin is seen when the vessels are constricted or of moderate tone, and the blood flow normal or rapid. A slowed cutaneous blood flow and constricted superficial vessels tend to produce a leaden or ashen type of cyanosis for, as mentioned above, the dark blood in the deeper venous plexuses then becomes faintly visible.

The temperature of the skin depends largely upon the rate of blood flow through its vessels. The radiation of the body's heat is carried out to an important extent through the medium of the cutaneous vessels (ch. 53). The warmer blood of deeper regions is diverted through the cutaneous channels and becomes cooled in its passage. So the cyanotic skin (if due to reduced cutaneous blood flow) is usually cool, and the flushed, scarlet skin hot. But the skin may be pale with constricted superficial vessels and yet radiate a large quantity of heat, if the blood flow through deeper vessels is rapid. Sometimes such a pale skin is hotter than another of a redder color. In the former instance, heat from the swiftly flowing blood in the deeper vessels is simply conducted through the overlying skin and dissipated from the surface through radiation and convection.

Vascular responses of the skin to stimulation by mechanical and other agencies

THE WHITE REACTION If the surface of the skin be stroked lightly with a blunt "pointed" instrument, a line of pallor appears in from 15 to 20 seconds which traces the path taken by the instrument. The line attains its maximal intensity in from a half to 1 minute, and then gradually fades to disappear in from 3 to 5 minutes. This *white reaction* must not be confused with the white trail which follows *immediately* in the wake of the instrument and is due simply to the expression of blood from the vessels by the stroke. The white reaction proper is due to *direct stimulation of the capillary wall and has not a nervous basis*. It has been shown by Lewis to be due to the tension exerted upon the walls of the minute vessels—capillary loops, collecting venules and especially of the subpapillary venous plexus—which respond to the stimulus by contraction. The sharply delineated character of the white line, and the fact that it can be obtained after the circulation through the region has been occluded by compression of the larger vessels, show that it is an active capillary response, and not the result of arteriolar constriction. The contractile force exerted by the capillary wall (mainly of the metarterioles, and precapillary sphincters) is such that they are capable of resisting a pressure of from 50 to 100 mm Hg. In other words, these vessels can remain contracted in the face of any pressure that could normally reach them through the arterioles.

THE TRIPLE RESPONSE This comprises (1) *the red reaction*, (2) *the flare*, and (3) *the wheal*.

(1) *The red reaction* If the pointed instrument be drawn more firmly across the skin, especially of the forearm or back, a red instead of a white band appears after a somewhat shorter latent period (3 to 15 seconds), reaches its maximum in from a half to one minute, and then gradually fades. The time of its disappearance is variable, this may occur in a few minutes or be postponed for half an hour or more. It can be seen to assume a bluish tinge before it fades. Like the white reaction it is strictly localized to the line of stroke, it is due to *dilatation* of the capillary vessels. The red reaction can be induced in its full intensity in the skin from which the circulation has been occluded, so it is due to active dilatation of the latter vessels and not merely a passive result of arteriolar dilatation (see p. 309). By means of a thermocouple placed upon the red line a rise in

temperature indicating increased blood flow may sometimes be detected. *The red reaction is not dependent upon nervous mechanisms since it occurs after section and degeneration of the cutaneous nerves*.

Pale lines bordering the central red band can frequently be seen. These are due to capillary contraction resulting from tension exerted upon the skin on either side of the line of stroke. The white halo is identical in every way, e.g., latency, duration and causation, with the white reaction described above. A certain diagnostic importance, quite unjustified, has in the past been attached to the red reaction with pale borders. It was spoken of as the *tache cérébrale*, or as *Trousseau's phenomenon*, and was believed to signify the existence of certain meningeal or cerebral lesions. It is, however, a perfectly normal response to mechanical stimulation.

(2) *The spreading flush or flare* If the stimulus is unusually strong, or is repeated often enough, the reddening of the skin is not confined to the line of stroke but surrounds it for a variable distance (1 to 10 cm) according to the intensity of the injury inflicted. The temperature in the suffused area is definitely raised. This flare reaction appears a few seconds (15 to 30) after the local red line, and fades sooner but remains a bright arterial color throughout. It is due to dilatation of the arterioles, since it does not appear after the circulation of the part has been occluded by means of a tourniquet, also, unlike the red reaction, *the flare is dependent upon local nervous mechanisms (axon reflex)*. It occurs after the nerves are divided but not after they have degenerated.

(3) *Local edema or wheal* When the stimulus is still more intense, the skin along the line of the injury becomes blanched and raised above the surrounding area to a height of 1 or 2 mm or even more. Such a wheal or welt is commonly produced in a normal person by the lash of a whip and other types of strong localized stimulation. The circular wheals of an urticarial rash (hives) are similar in character. In susceptible individuals, even light stimulation, such as drawing a pencil with moderate pressure over the skin of the back, will produce linear wheals surrounded by a diffuse red halo along the pencil's track. In this way letters or other designs may be embossed upon the skin (fig. 284). This phenomenon, which is spoken of as *dermographism* or *factitious urticaria*, has been considered pathological, though actually it can be demonstrated



FIG. 28.4 Dermographism (from Adams's textbook of Pathology, after Hyde and Ormsby)

to a greater or less degree in many young and perfectly healthy individuals, and can be elicited in some degree by repeatedly stroking the skin of the back of almost any normal person.

The wheal is preceded by and completely replaces the usual red reaction. It makes its appearance in from 1 to 3 minutes from the time of injury and is at its maximum height in 3 or 5 minutes. It is surrounded by the flare described above. The raised patch at first is clearly cut, but as time passes it increases in diameter and decreases in height, loses its sharpness and finally, though perhaps not for some hours, disappears. The wheal is due to the transudation of fluid from the minute vessels involved previously in the red reaction, it is, therefore, a localized edema. The gradual reduction in height and sharpness of the wheal and its final disappearance are due to the diffusion of the exuded fluid into a wider area of skin. Sometimes the fluid collects beneath the horny layer of the skin and strips it from the underlying epidermis. Such a collection of transuded fluid constitutes a blister.

Increased permeability of the capillary wall is judged to be the immediate cause of the localized edema which constitutes the wheal. Increased intracapillary pressure, distension of the capillary lumen or reduction in pressure in extracapillary spaces have been shown by Lewis to be not responsible. Suction, for example, amounting to a negative pressure of 90 mm Hg, applied to the surface of the skin over the line of the red reaction or over the area of the flare does not cause a wheal to form. Nor will the application of a positive pressure of 50 mm Hg prevent its appearance. Increasing the intracapillary pressures by compressing the veins does not cause a wheal to form more readily along the track of a red reaction, as a matter of fact a wheal upon a congested area is less pronounced. That increased permeability rather than simply a

rise in the filtration pressure is the dominating feature is also manifest by the high protein content of the transuded fluid. This more nearly approaches that of blood serum than the fluid of ordinary edema.

Wheal production does not depend upon a nervous mechanism. Though an accompanying flare is not an essential feature of a wheal, since the phenomenon can be elicited in denervated skin, the red halo is nearly always present, and in normal skin the degree of whealing is usually proportional to the latter's intensity. When, for instance, the injury is inflicted while the circulation is occluded, so that the flare is prevented, the wheal does not appear until the circulation has been again restored. This simply means, however, that though increased capillary permeability has occurred during the period of occlusion, blood fluid is not available in sufficient quantity for the production of the edematous swelling.

H-substance

A considerable weight of indirect evidence has been presented by Lewis to support the conception that a diffusible substance is responsible for the three reactions comprising the triple response. This material, which he calls *H-substance*, is thought to be liberated by the injured cells of the epidermis lying beneath the horny layer and superficial to the papillae. The possibility that more deeply lying tissues, when subjected to injury, may release the substance is not excluded, but a needle which does not penetrate beyond the epidermis elicits the typical threefold reaction. The chemical substance closely resembles *histamine* in its action. It apparently causes the red reaction and the wheal by a direct action upon the capillary wall. The flare is believed to be due to chemical stimulation of the sensory endings of the skin, thus bringing about arteriolar dilatation through the mechanism of the axon reflex (p. 290).

Though, as just pointed out, the evidence for this theory is indirect and actual proof of the existence of such a humoral mechanism is lacking, the results of Lewis's experiments carry conviction. The reader is referred to this author's original papers or to his monograph for the details of the various ingenious experiments performed, and the clear arguments set forth in support of his thesis. The evidence which has been brought forward can be no more than briefly cited here. The following is a summary.

(1) Histamine (1/1000 histamine phosphate) when pricked into the skin reproduces the triple response.

characteristic of mechanical injury. The individual reactions, local red reaction, wheal and flare produced by histamine are indistinguishable in all their features from those resulting from the various types of skin injury, pricking, scratching, freezing, burning, etc.

(2) Histamine can be extracted from many tissues of the body including the cells of the epidermis.

(3) An individual in whom the cutaneous reactions were readily produced was chosen and the circulation to the two arms occluded by means of a pneumatic armlet. A minute after circulatory arrest a stroke was made upon one arm, and on the other 9 minutes later. A typical flare resulted in each instance when the circulation was restored to both arms simultaneously. If the flares were due purely to a nervous mechanism, they should both fade at the end of the same interval of time following the application of the stimulus. That is, the second flare should disappear 9 minutes after the first. As a matter of fact they disappeared almost simultaneously. This observation though inexplicable upon a nervous basis conforms with the theory that the responsible agent is a chemical substance released by the injury. It also indicates that the chemical remains at the site of injury, its effect subsiding only after the blood flow to the skin has been restored. It is clear that if the flare is due to a released substance, then the same period of time will be required for the blood to wash away the material from the stimulated point in each arm. Identical results were obtained when, instead of a mechanical stimulus, histamine itself was pricked into the skin (fig. 28.5).

(4) Two wheals with their surrounding flares were made to appear upon the skin of the arm, one a short distance above the other. An elastic bandage was then wrapped tightly around the arm so as to cover the lower wheal and the lower half of the corresponding flare. The lower half of this flare was in this way deprived of its circulation while its upper half was visible above the bandage. The entire circumference of the upper flare was beyond the margin of the bandage. The bandage was retained in position for 20 minutes. The upper flare was found to fade gradually and to disappear in the usual length of time. The visible part of the lower flare, however, retained its vivid hue throughout the period of occlusion and only upon removal of the bandage and the re-entry of blood into the lower half of the flare did fading commence. This result is interpreted as indicating that the released substance being retained in the occluded area exerted a persistent effect upon the nerve endings there, and through axon reflexes sustained the arteriolar dilatation in the unoccluded half.

(5) Lewis and Harmer found that when wheals were induced by stroking an extensive area of the back of a susceptible subject, generalized circulatory responses similar to those resulting from the hypodermic administration of histamine, namely, flushing of the face and other parts of the body surface, and a distinct fall

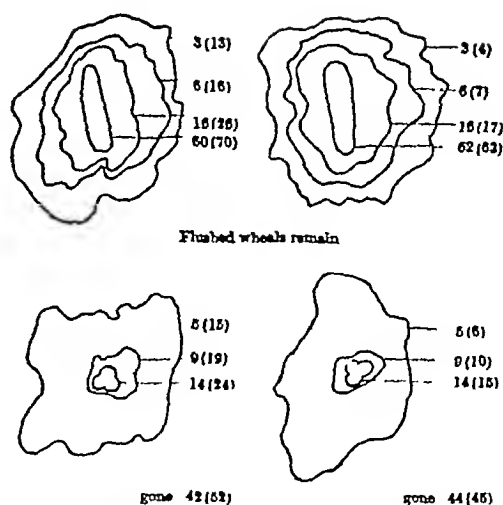


FIG. 28.5 Reactions caused by histamine and by injury (upper figure) in an urticarial subject. The vessels of the two arms were occluded. One minute later the left forearm was stroked (above) and a group of three histamine punctures (1 in 3,000 dilution) was laid down (below), 9 minutes later the right forearm was similarly and symmetrically treated, the vessels of both arms were released 10 minutes later, i.e., at the 11th minute. The numbers outside the brackets indicate the times in minutes at which the contours were outlined after restoration of the circulation. The numbers within the brackets give the times elapsing after application of the stimuli. In the stroke figures the fading was followed until the redness became confined to the wheals. (After Lewis.)

in blood pressure resulted. The fluid drawn from urticarial wheals was also tested for a histamine-like action. The results, however, were not conclusive. Only in some instances could such an action be demonstrated. The fluid was also tested upon the guinea-pig's uterus, but again little light was thrown upon this question, for, though the fluid stimulated this organ to contraction, normal plasma caused a similar effect. More recently it has been shown that extensive stimulation of the skin of a susceptible person causes increased secretion of gastric juice. This is a well-known histamine effect (pp. 295 and 518).

(6) A person upon whose skin urticarial wheals can be provoked with unusual facility by stroking (dermographism) shows a sensitivity to histamine injections that is no higher than normal. This fact indicates that the susceptibility of such an individual does not depend upon a greater responsiveness of his capillary vessels. The greater readiness with which his skin reacts to mechanical stimulation is therefore attributed to an exaggerated sensitivity of the cells of the skin to injury and a more prompt release of a diffusible substance.

The cutaneous reactions following injuries of various sorts, e.g., burning, freezing, electrical stimulation and the more slowly-developed erythema resulting from ultra-violet irradiation, are believed to be produced by the same humoral

mechanism. The cardinal signs of inflammation—redness, heat, and to some extent the swelling and pain—can be similarly explained, namely, direct action of the dilator H-substance upon the minute vessels, and an indirect effect through the medium of the axon reflex, upon the arterioles of the surrounding area.

Lewis has extended this conception to embrace the cutaneous vasodilator effects of antidromic impulses discharged under ordinary circumstances from the central nervous system through the posterior roots (p. 276). The nervous impulse, it is conceived, brings about its effect upon the vessels through the release of H-substance from the cells of the epidermis, and not by a direct action upon the vascular wall.

The cutaneous lesions characteristic of herpes zoster are ascribed to the liberation of H-substance from the epidermal cells by antidromic impulses. Head and Campbell have demonstrated inflammatory changes, hemorrhages and destruction of nerve cells in the posterior root ganglia. Similar skin changes may follow involvement of the roots by malignant disease or injuries, such as the so-called trophic ulcers accompanying lesions of the sensory side of the nervous system, e.g., along the course of lightning pains in tabes. The erythema and blistering of the skin induced by the suggestion of burning or other injury during the hypnotic state may possibly be produced in a similar manner. Also it is probable that many other types of cutaneous lesions familiar to the dermatologist, e.g., vesicles, urticaria, maculae, purpuric spots, etc., have a neuro-chemical basis.

It has been suggested by Dale that acetylcholine (ch. 72) probably also plays a rôle in cutaneous vascular reactions. It is not unlikely that this substance is liberated at the nerve ending on the arteriole as a result of impulses either discharged antidromically from the central nervous system or arising peripherally and reaching the vessel along the path of an axon reflex. In the latter instance the cutaneous stimulus is conceived as causing the liberation of H-substance. This excites cutaneous nerve terminals, and the impulses thereby set up being transmitted to the nerve endings on the arteriole, cause the release of acetylcholine. This conception would, therefore, hold acetylcholine directly responsible for the arteriolar dilatation characteristic of the flare

REACTIVE HYPEREMIA AND BIER'S SPOTS. If the circulation to a part is arrested for a time (as by com-

pression with a tourniquet or a sphygmomanometer armlet) and then released, the skin flushes, the volume of the part increases and the blood flow through it is greater than before the vessels were occluded. The hyperemic reaction occurs though all nerves supplying the part have been severed, and it has been shown by Lewis and Grant that it occurs independently of any direct nervous influence (e.g., axon reflexes). These observers furnish evidence that it is due to the accumulation of H-substance in the skin during the period of circulatory arrest. It has also been demonstrated by Barsoum and Smirk that the concentration of a substance with the biological properties of histamine increases in the venous blood of the arm after a period (10 to 30 minutes) of circulatory arrest.

During the period of circulatory arrest the skin becomes mottled, blanched areas (Bier's spots) and bright blotches appearing on the background of cyanosis. The bright areas are caused by blood delivered through the nutrient arteries of the bones and anastomatic channels to the main vessels of the limb below the obstruction. This trickle of blood forces oxygenated blood from the occluded arteries into the cutaneous capillaries. The pallid areas, which were first described by Bier in 1898, are due to capillary contraction. A new significance has been given to the latter by the observations of Rous and Gilding, who found that they occur also in the skin and other non-vital tissues of animals when the blood volume is considerably reduced. These authors look upon the localized contraction of the vessels in such instances as part of the mechanism whereby the capacity of the circulatory system is reduced and the blood supply to vital tissues thus maintained at the expense of those of less importance.

CAPILLARY PULSATIONS

Rhythmical changes in skin color—flushing alternating with pallor corresponding to the heart beats—may occur when the smaller arterioles of the skin become dilated. It is seen in aortic regurgitation and in arteriovenous aneurysm, and is usually ascribed to the high pulse pressure characteristic of these conditions. Lewis, however, has shown that high pulse pressure, though it enhances the phenomenon, is not a necessary condition for its appearance, for it occurs also when the pulse pressure is low (see p. 147). The essential factor concerned in capillary pulsation is a certain degree of dilatation of the terminal arterioles of the skin. The opening of these vessels reduces the peripheral resistance and the flow in the capillary then tends to become intermittent instead of continuous (p. 149). Pulsation in the capillaries may be demonstrated in most normal

young persons, especially if peripheral vasodilatation be induced by immersing the hand in hot water, or by means of a vasodilator drug, and then observing the skin through a glass slide pressed lightly upon it. The pulsation is seen particularly well in the more highly vascularized regions, e.g., the finger tips, the cheeks, lips or ear lobes, or on the outskirts of inflamed regions. It may frequently be demonstrated in certain of these regions in normal individuals, even without preliminary dilatation, and is especially evident under the low power of the microscope. In elderly persons, however, it may be impossible of demonstration, even after the application of heat, and then it is assumed that the vessels are incapable of the necessary degree of dilatation. The capillary loops, the collecting venules and the subpapillary venous plexus, i.e., those vessels responsible for skin color, are involved in the pulsation. Capillary pulsation cannot therefore be considered as an unequivocal indication of aortic regurgitation or other pathological condition.

CAPILLARY PRESSURES

The blood pressure in the arterial and venous limbs of the capillary loops of the human finger has been measured by Landis. A fine cannula was inserted into either limb of the loop as desired and the pressure measured directly by means

of the apparatus shown in fig 28.6. With the hand at or above the level of the auricle (level of the manubrium sterni) the capillary pressure remains constant but increases when the hand is moved to a lower level. The increase in pressure is in direct proportion to the distance between the new level and the base of the heart and is due to the hydrostatic pressure of the column of blood in the veins (p. 169). The average pressure in the arterial limb of the loop was found to be 32 mm Hg, and in the venous limb 12 mm Hg when the hand was at the level of the auricle. In the arterial limb the blood pressure was higher, in the venous limb lower, than the osmotic pressure of the plasma (see p. 31). The application of heat, venous congestion or the production of a histamine flare elevated the capillary pressures well above these values. The application of cold caused first a fall in pressure and then a small rise above the original pressure. The production of a wheal, as by freezing, resulted in a rise of the average pressure, in the arterial limb to 49 mm Hg, and in the venous limb to 32 mm Hg. In Raynaud's disease the pressure in the capillary loop is greatly reduced during the vascular spasm. (See table 25.)

ARTERIO-VENOUS ANASTOMOSES

Grant and his associates have shown that in the ear of the rabbit, communications (20 to 70 micra

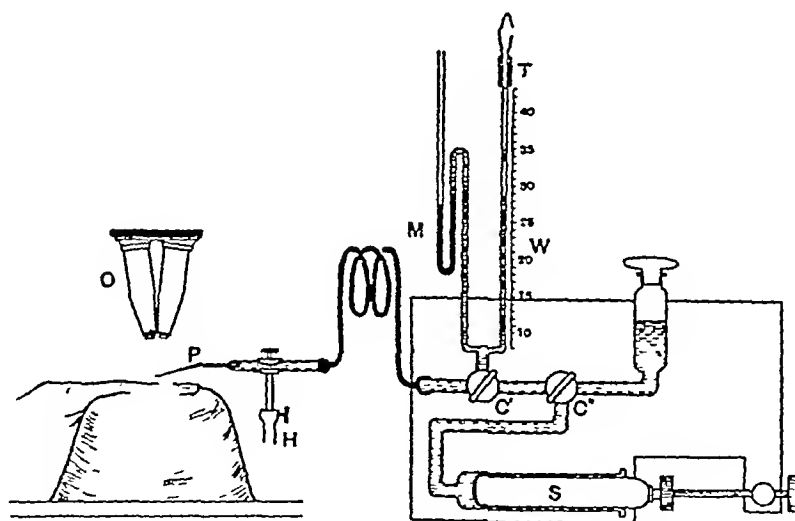


FIG. 28.6. Diagram of apparatus for the determination of capillary pressure by micro-injection. C', C'', three-way stopcocks, H, holder of micromanipulator, M, mercury manometer, O, binocular microscope, S, syringe. The system is filled with physiological saline to which has been added 0.3 per cent sodium citrate as an anticoagulant. When the tubing (W) is closed at T the height of the mercury column indicates the pressure exerted upon the contents of the micropipette inserted into the capillary. By means of the syringe, the pressure in the system is adjusted until the corpuscles remain in the same general position in the tip of the pipette, oscillating back and forth with each heart beat. The reading of the manometer then gives the capillary pressure. (After Landis, *Heart*, 1930, 15, 209.)

TABLE 25*

Grad ent of pressure in the capillary loop of skin of hand

PRESSURE MEASURED IN	NUMBER OF OBSERVATIONS	CAPILLARY PRESSURE	
		Range	Average
Arterio-venous limb	125	21-48	32
Summit of loop	29	15-32	20
Venous limb	99	6-18	12

* Modified from Landis.

in diameter) exist normally between the smaller arteries or arterioles and the corresponding venous channels (fig 287), through which the blood may be shunted and capillary areas short circuited. Shunts have been observed in the omentum of the dog, in the rat's mesoappendix and in the human skin. They are also present in the toes of birds. Similar shunts connecting metarterioles with venules have been described by Zweifach, their walls contain a considerable amount of smooth muscle. At ordinary room temperature these anastomotic channels are closed, but they open when the air temperature rises or falls below a certain critical level. Local heating of the rabbit's ear does not cause dilatation until the tempera-

ture rises to between 35° and 40°C. Local cooling causes dilatation which commences when the temperature is reduced to below 15°C, above this temperature the capillaries are constricted. At 5°C the dilatation of the anastomotic channels is general. This reaction is brought about reflexly and is not confined to the portion of the ear actually cooled, but involves the whole member and even the ear of the opposite side. In man, the blood flow in the vessels of the terminal phalanges of the fingers may increase some 100 times as a result of the reflex dilatation of arterio-venous anastomoses. The communicating channels are believed to have two important functions, (a) regulation of the body's temperature by increasing the radiation of heat, since when dilated they permit an enormous quantity of blood to flow through the peripheral parts. The importance of this mechanism is evident from the fact that in the rabbit the body temperature can be raised or lowered at will by heating or cooling the ears. (b) Maintenance of the temperature of outlying parts of the body against local cooling.

A-V anastomoses are found in tissues other than the skin, e.g., the lungs and myocardium.

The arterio-venous anastomoses are supplied by constrictor sympathetic nerves, and are constricted by adrenaline. They are dilated by mechanical stimuli, by the stimulation of sensory nerves and by histamine or acetylcholine. The reactions are therefore essentially the same as those of the arterioles. In man, similar channels have been demonstrated connecting the arterioles and veins of the terminal phalanges.

THE CORONARY CIRCULATION

The heart muscle is supplied with blood from the two coronary arteries, right and left, which arise directly from the aorta close to its origin. The *left coronary artery* is a short trunk which runs forward between the root of the pulmonary artery and the right auricular appendix to the upper end of the anterior interventricular groove (sulcus). Here it divides into the *left circumflex*, which runs in the coronary sulcus to the left margin of the heart, and the *interentricular* (ramus descendens), which follows the anterior interventricular groove to the apex of the heart. The *right coronary artery* runs forward between the pulmonary artery and the right auricular appendix to the coronary sulcus, then downwards and to the right to reach the right extremity of the inferior cardiac margin. It then curves to the left in the coronary sulcus to the posterior end of the inferior interventricular groove, where it gives off its *interentricular* branch. The latter

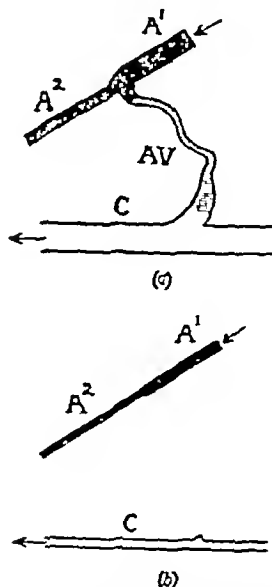


FIG 287 Reaction of anastomosis and associated vessels to lowering of body temperature. A1, artery, A2, arteriole, AV, arterio-venous anastomosis, C, vein. Upper figure (a), anastomosis open, lower figure (b), anastomosis closed. (After Grant.)

runs forward in the inferior interventricular groove to the apex, anastomosing here with the corresponding branch of the left coronary. After giving off this branch the right coronary ends by anastomosing with the circumflex branch of the left coronary artery. In the dog 80 per cent of the blood supply to the heart is carried by the left coronary artery (50 per cent by the circumflex branch, and 30 per cent by the anterior interventricular (descending) branch).

The branches of the A-V bundle are supplied by twigs from both right and left coronary arteries.

Though there is considerable variation in the distribution of the two coronary vessels, in the most common arrangement (48 per cent of human hearts according to Schlesinger), the greater part of the heart receives branches from the right coronary artery which supplies all of the right ventricle, the posterior half of the interventricular septum and part of the left ventricle, a special branch of the right coronary goes in the majority of cases to the sino-auricular node, another to the auriculo-ventricular node and bundle.

In a smaller proportion of instances (34 per cent), the distribution of each coronary vessel is largely confined to the corresponding ventricle, thus, the right ventricle and the posterior part of the interventricular septum is supplied by the right vessel, while the left ventricle and the anterior part of the septum is supplied by the left. In 18 per cent of the 225 hearts examined by Schlesinger, the distribution of the *left coronary* artery predominated. It supplied the whole of the left ventricle and a variable portion of the right chamber and of the septum.

Schlesinger has analyzed his anatomical findings and correlated them with the incidence of coronary disease. His results suggest that the second group, in which the blood supply to the heart is not predominantly through either the right or the left vessel, is the least vulnerable to arteriosclerotic changes and their effects. The group in which the distribution of the left vessel predominates appears to be the most susceptible, while that in which the right vessel supplies the greater part of the heart stands in an intermediate position with regard to the incidence of coronary disease.

The return of the blood from the myocardium to the right side of the heart is through a *superficial* and a *deep* venous system. The former, which lies just beneath the epicardium, consists mainly of (1) the *coronary sinus*, situated in the lower part of the coronary sulcus between the left auricle and the corresponding ventricle. It opens into the posterior part of the right auricle behind the right auriculoventricular orifice, and drains the greater part of the blood from the left coronary artery, as well as part of the blood of the right coronary, (2) the *great cardiac vein*, which ascends

alongside the anterior interventricular branch of the left coronary artery and, turning to the left in the coronary sulcus, empties into the coronary sinus, and (3) the *anterior cardiac* veins, usually 2 or 3 in number (2 to 5 in the dog), seen on the surface of the right ventricle. In the dog from 50 to 92 per cent of the blood of the right coronary artery drains into these latter vessels (Gregg and associates). The left coronary artery contributes a variable amount of blood to the anterior cardiac veins, but the proportion is considerably less than that derived from the right artery. The anterior cardiac veins empty into the right auricle a little above the border of the right A-V valve.

The deep venous system consists of a number of small veins which arise in the substance of the myocardium and empty into the right chambers by the Thebesian and other luminal vessels (see also pp. 322-323).

The heart muscle possesses an extraordinarily rich capillary supply. The ventricular muscle is much more vascular than the auricular walls or than the auriculoventricular bundle. In the rabbit's ventricles the combined capillary length per square cm. of tissue is about 11 meters as compared with 6 meters for active skeletal muscle. At birth a single capillary supplies 4 or 5 cardiac fibers. During growth the muscle fibers increase in size but not in number and each fiber in the adult heart receives a capillary vessel (Wearn), the number of vessels per unit area of tissue remaining unchanged from that at birth. In cardiac hypertrophy, on the contrary, the enlargement of the individual fibers is not accompanied by a corresponding increase in the number of capillaries. That is, the fiber-capillary ratio remains unchanged, but the number of vessels per unit area is reduced. The hypertrophied heart, therefore, suffers a relative reduction in its blood supply.

ANASTOMOSES

The coronary arteries form three types of anastomoses or communications.

(1) Potential anastomoses between branches of one coronary with those of the other. But in the great majority of normal human hearts, the coronary branches are essentially "end arteries" (see p. 323), that is, no pervious channels of importance exist between them.

(2) Communications with the cavities of the heart.

(a) The THEBESIAN VESSELS, described by

Thebesius in 1708, are small venous channels which open into the auricular and ventricular cavities. Thebesius demonstrated the communication of these vessels with the coronary system by inserting a blow pipe into a coronary artery and blowing through it after immersing the heart in fluid; bubbles were observed issuing from the heart cavities. The relations of these channels have been studied within recent years in the human heart by Wearn, and by Grant and Viko in the heart of the sheep. They have been shown to communicate on the one hand with the cavities of both ventricles (mostly the right) and, on the other with the coronary sinus and other large coronary veins, and hence with the capillary bed of the ventricular muscle.

There are also many channels connecting the cardiac veins with one another, and with extracardiac veins tributary to the superior and inferior vena cavae, arteriovenous anastomoses between smaller vessels are fairly plentiful in the myocardium.

(b) **DIRECT COMMUNICATIONS BETWEEN THE CORONARY ARTERIES AND THE VENTRICULAR CAVITIES, I.E., COMMUNICATIONS OTHER THAN THROUGH THE CAPILLARY BED.** These channels, first recognized by Vieussens in 1706, have been clearly demonstrated by Wearn and his associates in human hearts and are of two types which they term *arterio luminal* and *arterio sinusoidal* respectively. The arterio luminal vessels are coronary branches from 0.2 to 1.0 mm in diameter. They run a fairly direct course and usually retain their

arterial characters up to the point where they open into the lumen of the heart. The arterio-sinusoidal vessels are branches of a coronary artery which break up into a number of irregular channels varying from 50 to 250 micra in diameter. These latter are called *myocardial sinusoids*. The walls of the sinusoids are composed of a single layer of endothelial cells, run a meandering course, and anastomose very freely with one another. The arterio sinusoidal vessels thus provide, through the myocardial sinusoids, a means of free communication between the ventricular cavities and the coronary arteries.

(3) **EXTRACARDIAC ANASTOMOSES.** Potential communications between auricular twigs of the coronary arteries on the one hand, and branches of the internal mammary and aorta—pericardial, bronchial, phrenic and esophageal—on the other, have been demonstrated by Hudson, Montz and Wearn. The site of anastomosis between these two systems is in the pericardial fat, and around the openings of the great vessels. These observers injected the coronaries from the aorta with a suspension of lamp black in gum acacia solution. A rich network of injected vessels appeared over the entire surface of the parietal pericardium (fig 28.8), over the diaphragm and in the walls of the pulmonary vessels (*vasa vasorum*). The trachea, esophagus and mediastinum also showed injected vessels. Upon injection of the aorta, after tying it off just distal to the orifices of the coronaries, and at the level of the diaphragm, the injected material was found in the coronary vessels around the openings of the pulmonary veins. Pericardial adhesions increase the extracardiac anastomoses.

Beck and Tichy, in a series of experiments upon dogs, demonstrated the feasibility of developing the extracardiac anastomoses to such an extent that they could sustain the nourishment of the heart after almost complete occlusion (85%) of both coronary vessels. Adhesions were produced by incising the myocardium and stitching the pericardium into the wound, or by removing the epicardium and the endothelium of the pericardium. In other experiments an adjacent thoracic muscle, or the omentum drawn through an incision in the diaphragm, was utilized to establish communication between the coronary vessels and an extracardiac vascular bed. Occlusion of the coronary arteries was brought about gradually by means of silver clips which were tightened in successive operations.² An operation devised on the basis of his ex-

² Occlusion of the coronary sinus in animals is also claimed to encourage the development of collateral channels and to reduce the mortality following ligation of a large coronary branch.



FIG 28.8 Drawing of the inner surface of the parietal pericardium and the inferior surface of the diaphragm. The coronary arteries were injected and the heart removed to expose the sites of anastomoses between coronary and extracardiac vessels. (After Hudson, Montz and Wearn.)

perimental work has been performed by Beck upon patients suffering from angina pectoris due to coronary disease, the pectoralis major muscle was employed

O'Shaughnessy has employed an omental graft (*cardio-omental cardiopexy*) with success in the treatment of angina. The omentum possesses the power of vascularization above all other tissues. The effectiveness of such a graft was clearly demonstrated in experiments upon greyhounds. Animals with the descending branch of the left coronary ligated and with an omental graft were able to race over 500 yards.

Heinbecker and Barton injected an irritant mixture into the pericardial sac of dogs. Through the adhesions set up, effective collateral channels are said to have been established.

Interesting and encouraging as are these various procedures designed to increase extracardiac anastomosis, their value in angina pectoris has not been proved (see Marvin).

It does not appear that anastomotic communications play any significant rôle in supplying blood to the *normal* heart. When a large coronary branch is suddenly occluded, the back-flow in the section of the vessel distal to the occlusion is only about 1 cc per minute and the myocardium normally supplied by the obstructed vessel ceases to contract within from 1 to 2 minutes. Nor does occlusion of other coronaries reduce the back-flow in the distal section of the previously occluded vessel. On the other hand, gradual occlusion of a coronary is followed by the development of a back-flow from other coronary branches as a result of altered pressure gradients. The pressure in the peripheral section of the occluded vessel increases and may approach the aortic pressure. In the dog, the retrograde flow usually amounts to from 30–40 cc per minute and, in some instances, is over 100 cc per minute. The myocardium, to which the occluded vessel is distributed, shows normal contractions (Gregg and associates). The importance of anastomatic channels in supplying the heart muscle in coronary disease is indicated by the observations of Wearn, he has reported cases in which postmortem examination revealed the main coronary arteries to be completely occluded, yet the coronary occlusion had not been the cause of death, but must have been of long standing. The heart muscle, apparently, had in these instances been nourished through channels communicating with the ventricles or through extra-cardiac anastomoses, or through both of these routes. According to Gross, anastomoses between the branches of the coronaries multiply with age and the vascularity of the

muscle, especially of the left ventricle and septum, increases.

The possible functional significance of the communications between the coronary system and the ventricular cavity is indicated by the experiments performed some years ago by Pratt. He was able to keep the *isolated* heart of the cat beating for an hour by perfusing it through a cannula tied into the ventricle. Katz and his colleagues introduced a pure culture of killed bacteria into the superior vena cava of a heart-lung preparation (precautions having been taken that bacteria did not enter the coronary system from the arterial side). The bacteria were found later in the sinusoidal spaces, capillaries and the small arteries of the myocardium. These authors conclude that though the quantity of blood conveyed from the ventricles to the *normal* heart muscle is small, in some pathological states associated with gradual narrowing or occlusion of coronary vessels, it may be sufficient to aid significantly in nourishing the myocardium. The results of Stella's experiments on the heart-lung preparation of the dog also indicate that in a heart whose muscle is well supplied with blood through the coronary system, back-flow does not occur to any important extent from the ventricles into the coronary vessels. Even when the ventricular pressure was higher than the coronary pressure no evidence of back-flow was secured.

CORONARY CIRCULATION TIME AND VOLUME FLOW

The coronary circulation is one of the shortest in the body. Evans and Starling calculated that 60 per cent of the blood of the coronary circulation is delivered into the right auricle by the coronary sinus. The remaining 40 per cent must find an outlet through other channels—the Thebesian and other communications mentioned above.³ According to Wearn, the proportion of blood returned otherwise than through the coronary sinus may, under certain conditions, be much greater than 40 per cent. The blood returned through the coronary sinus is not a constant proportion of the total amount, but varies under different conditions. The blood drained through the coronary sinus comes mainly from the left coronary artery (64 to 83 per cent), and over 70 per cent of the blood carried by this artery is returned to the auricle through the sinus. Clamping

³ The latter is probably only a very approximate figure which varies considerably under different conditions, such as, especially, the pressure in the right ventricle.

the left coronary artery, therefore, reduces very greatly the outflow from this venous channel. The anterior cardiac veins, Thebesian vessels and other luminal vessels drain regions of the heart muscle supplied by the right coronary artery, ordinarily, only a small part of the right coronary blood drains into the coronary sinus. A rise of pressure in the pulmonary circulation (and therefore of the pressure in the right ventricle) reduces the pressure gradient between the coronary arteries and the Thebesian vessels and causes thereby, slowing of the flow along these channels. Such slowing interferes very seriously with the nutrition of the right ventricular muscle. Moe and Visscher found that a reduction in the aortic pulmonary pressure difference to 40 mm Hg, either by a fall in aortic pressure or a rise in pulmonary pressure, diminished the Thebesian flow to 20 per cent of the normal. The partial dependence of the nutrition of the right ventricle upon the venous return through the Thebesian vessels is suggested as the probable reason that this side of the heart can work for a relatively short time against a high pulmonary resistance (e.g. mitral stenosis or pulmonary fibrosis) whereas the left ventricle is able to work for years against a high aortic pressure, since the flow from the coronary sinus is not affected by the high pressure in the left ventricle. A rise in aortic pressure by increasing the pressure gradient might therefore be expected to benefit cases of cardiac disease with raised pulmonary pressure. Such benefit has been observed.

The coronary circulation time, that is, the time taken for a red cell (following the shortest route) to pass through the coronary system, is about 8 seconds as compared with 20 seconds through the vessels of the limbs.

Experiments upon animals (heart lung preparation) indicate that about 5 per cent of the total output of the heart flows through the coronary system. The output of the human heart under basal conditions is between 5 and 6 liters per minute, but during strenuous muscular exercise it may, in the case of a large muscular subject, rise to 37 liters. In such an instance, if the volume of the coronary flow were 5 per cent of the total blood flow, nearly two liters would pass through the coronary system.⁴ The coronary flow

in man can be estimated in another way. The work of the heart, for example, can be calculated approximately from the cardiac output and the mean arterial blood pressure (p. 141). Taking the efficiency of the heart at 20 per cent, the oxygen consumed by the myocardium can be calculated from the work performed. The results of such calculations indicate that the heart muscle during strenuous exertion consumes as much oxygen as does the entire body during rest (Hill), or about 250 cc per minute. Assuming that during maximum muscular effort the coefficient of oxygen utilization is 0.7 (p. 374), that is to say, the heart muscle removes about 13 cc of oxygen from each 100 cc of blood (arteriovenous oxygen difference = 13 vols), then the delivery of 250 cc of oxygen requires a flow of $((250/13) \times 100 =)$ about 2 liters of blood through the coronary system per minute, or an amount of blood having a weight more than 6 times greater than that of the heart itself (300 gm). The coefficient of oxygen utilization of the heart muscle is probably higher than this during strenuous exercise which would give a lower figure for an equivalent oxygen consumption. The A-V oxygen difference of the coronary blood of the dog at rest is from 8 to 16 volumes per cent.

THE CORONARY FLOW DURING DIFFERENT PHASES OF THE CARDIAC CYCLE

Scaramucci, an Italian of the 17th century, was the first to recognize that unlike other arteries the flow through the coronary arteries occurred mainly during diastole, and that a reduced amount of blood entered the vessels during systole owing to their compression by the contracting cardiac muscle. Stroem offered an anatomical reason for the reduced flow during systole, namely, that the leaves of the aortic valves blocked the coronary orifices during this phase of the cycle. This explanation was soon shown to be incorrect, for the edges of the open valves in many cases do not reach as far as the coronary orifices. Langendorff studied the problem intensively in the perfused isolated heart, and our knowledge of the coronary flow during the cardiac cycle is based largely upon his work. In more recent times, Anrep and his colleagues in England, and Wiggers, Green and Gregg with their associates in America have been the outstanding contributors in this field.

The coronary inflow The flow in the coronary

⁴ This calculation probably gives too high a value, for it has been found by Eckenhoﬀ and associates that the percentage for the coronary flow varies inversely with the cardiac output.

system is affected profoundly by the compressing force of the cardiac muscle during systole (extra-vascular pressure). The two main factors affecting coronary inflow are, therefore, aortic pressure and extravascular compression. The flow through the peripheral vessels, which are surrounded by muscle and are directly compressed, is referred to as the *intramural flow*. The flow through the larger, more superficial coronary branches, is called the *extramural flow*, and is affected both directly and indirectly by compression during systole. The total flow (intra- and extramural) in a coronary vessel can be determined by perfusing the vessel through its orifice and measuring, by some form of flow meter, the quantity of fluid which enters. Wiggers and his associates measured the extramural flow by means of an optically recording flow meter in a coronary orifice, and calculated the intramural flow from pressure differences between the central and peripheral ends of a coronary branch. They found that the pressure in the peripheral coronary vessels never rises as high as the aortic pressure, that is, the pressure gradient is not abolished at any time during the cardiac cycle. The *intramural flow*, therefore, is not completely arrested. But, two sharp reductions in the flow occur, one in the isometric period and at the beginning of ejection, the other in the latter part of the ejection period. The rate of flow is greatest in mid-systole and throughout diastole.

The curve of extramural flow as measured at a coronary orifice, resembles that of the intramural flow, except that it is *completely* arrested during early systole and is greater during mid-systole and the first half of diastole. The arrest of the orifice flow in *early systole* is attributed to compression of the deeply placed vessels which forces blood backwards into the larger superficial vessels and thus opposes the flow from the aorta at this time. But as mentioned above the pressure in the peripheral vessels never rises as high as the aortic pressure, so that the onward flow in these vessels is never completely stopped. As the aortic pressure rises during systole, the superficial vessels are distended to accommodate more blood, the extramural flow is then increased over that in the deeper-lying and compressed intramural vessels. During early diastole, the peripheral vessels are quickly released from compression and the intramural flow increased. The flow through the extramural vessels, no longer impeded by back-flow, is augmented.

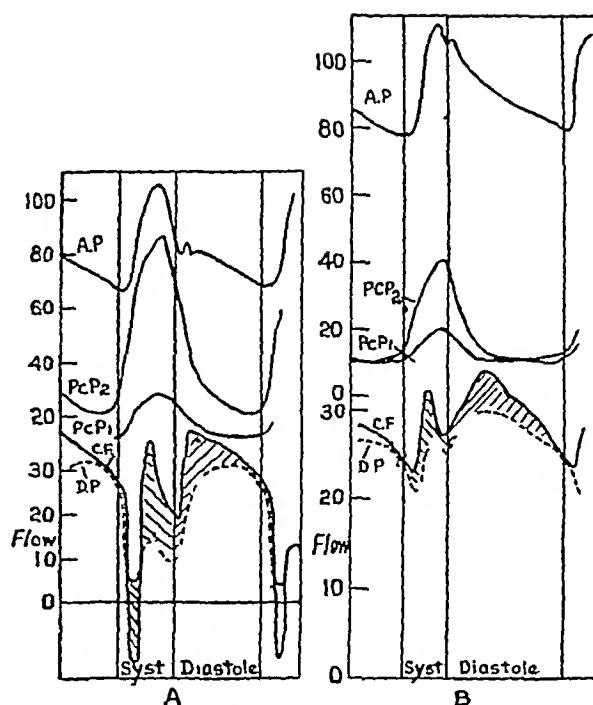


FIG 28.9 Comparison of the curves representing intramural and total inflow at a coronary orifice in the left coronary (A) and in the right coronary (B) together with the pressure curves from which the intramural flow was reconstructed. A.P., aortic pressure, PCP, peripheral coronary pressure as recorded, PCP₂, peripheral coronary pressure with proper ordinate value, CF (heavy line), inflow at a coronary orifice, DP (dashed line), differential pressure curve representing intramural flow. Shaded areas represent extramural flow. Ordinates, top, mm of mercury, bottom, flow in cubic centimeters per minute. (After Gregg "Blood Heart and Circulation", p 81 Science Press, 1940)

The contours of the blood flow curves of the *right coronary* differ from those of the left vessel in that the reduction in the rate of flow in early systole is not nearly so pronounced and the rate of total orifice flow does not exceed the intramural flow to as great a degree (fig 28.9). The flow during systole may approach or even exceed that during diastole.

The thermo-stromuhr of Rein (p 175), as modified by Baldes and Herrick, has been employed by Essex and associates for measuring the coronary flow in conscious animals (dogs). The chest was opened and the instrument placed in position. The connecting wires were then led to the exterior and the chest closed. After recovery from the operation, blood flow determinations were made from time to time under various physiological conditions. The flow was found to be increased during digestion by over 80 per cent and by about 400 per cent when the animal performed strenuous work.

Coronary pressures

The pressure curve of a coronary artery has a form closely resembling the aortic pressure curve. The minor differences which are to be seen in the pressure pulse of a coronary artery are not peculiar to that vessel but appear in pulses recorded from other branches of the aorta. The pressure in the right coronary artery is lower than in the left coronary but is considerably higher than the pressure in the right ventricle, whereas the pressure in the left coronary is lower than the left ventricular pressure. The pressure in the coronary sinus amounts to 10 or 12 mm Hg.

The outflow from the coronary sinus The flow from the coronary sinus is increased slightly during the latter half of auricular systole and very greatly during ventricular ejection. A graphic record of the outflow, therefore, shows two well-defined waves (cf fig 28 10). The first and smaller wave is probably due to the compression of the walls of the coronary sinus by the auricular muscle and the expulsion of a small quantity of blood. Evidence for the dependence of this wave upon auricular contraction is, that it occurs during the latter part of the P deflection of the electrocardiogram, that it disappears during auricular fibrillation, and that in heart block it synchronizes with the auricular contractions.

The large wave occurring during the ejection phase of ventricular systole is due to the blood being squeezed from the coronary veins by the

grip of the ventricular muscle. At the end of ventricular systole the outflow falls suddenly as a result of the release of the veins from the compressing force of the muscle. The vessels having been practically emptied must refill before the flow from the coronary sinus is resumed.

A diminutive wave is seen in the figure between the two waves just described. It occurs during the isometric period of ventricular systole but its cause is uncertain.

In coronary perfusion experiments, raising the pressure within the right ventricle increases the outflow from the coronary sinus, due to the diversion of blood from the Thebesian drainage system. If the perfusion pressure is quite low, and the pressure in the right chamber high, the flow in the Thebesian vessels may be reversed, and blood enters these from the ventricle, so that the outflow through the coronary sinus may actually exceed the coronary inflow.

FACTORS REGULATING THE CORONARY CIRCULATION

(1) **THE AORTIC BLOOD PRESSURE** The height of the mean aortic pressure, and especially of the mean pressure in the aorta following the closure of the semilunar valves (postdiastolic pressure), and the resistance in the peripheral coronary vessels are the most important factors determining the coronary inflow. For example, in aortic regurgitation, or arterio venous aneurysm produced experimentally, the coronary flow is decreased due to the sharp fall in diastolic pressure, but in these conditions during systole, the aortic pressure is elevated to a greater degree than the peripheral coronary pressure. The steeper pressure gradient results in a greater systolic flow which compensates, in part, for the decreased flow during diastole. The flow may be greater during systole than in diastole. In experimental aortic stenosis systole is prolonged and the coronary flow correspondingly reduced, the flow during diastole is not significantly affected.

(2) **INNERVATION OF THE CORONARIES** Woolard has demonstrated by histological studies that the coronary vessels are very richly supplied with both vagal and sympathetic fibers. The larger coronary branches were found to be innervated about equally by the two types of nerve, whereas the arteriolar innervation was mainly through the vagus, at least relatively few sympathetic fibers could be traced as far as the arterioles. The action which each of these two sets of fiber exerts upon

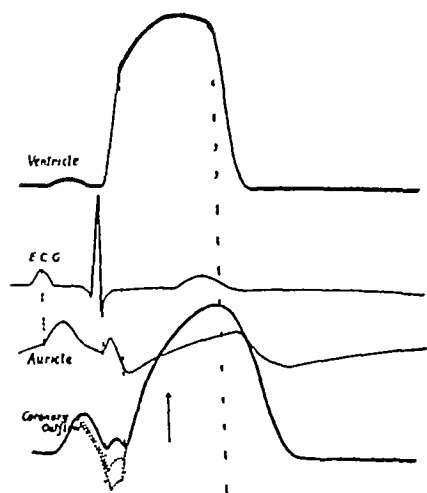


FIG 28 10 The relation of coronary outflow to other events in the cardiac cycle (After Anrep and associates)

the heart has, however, been a moot question for several years and a very difficult one for investigation, since stimulation of the vagus, or of the sympathetic, causes changes in the action of the heart which in turn influence the coronary circulation

Though the question of the actions of the vagus and sympathetic nerves upon the coronaries has not received a final answer, the weight of evidence indicates that, in the mammal, the vagus is vasoconstrictor, while the sympathetic and adrenaline are vasodilator. Anrep and Stacey observed an increase in coronary flow of 85 per cent after the administration of adrenaline, and several workers have reported an increased flow from stimulation of the stellate ganglia. Mann and associates, employing the thermo-stromuhr, observed an increase of 300 per cent in the coronary flow after the administration of adrenaline or ephedrine. Wiggers showed that stimulation of the peripheral end of the cut vagus in an animal under the influence of atropine, which in suitable dosage abolishes the vagal action upon the heart but not upon the coronaries, reduced the flow from a coronary vein. Diminution in coronary flow upon vagal stimulation was also demonstrated by Anrep and Segall, using the heart-lung preparation and the hot wire method of recording. The effect persisted though constancy of heart rate was secured by atropinization. Anrep and Segall also found that the coronary flow was increased considerably after vagal section, a fact which indicates that these nerves exert a tonic vasoconstrictor influence. The vagus, as we know, exerts its effect through the liberation of acetylcholine, and this substance has been demonstrated by Feldberg and Krayner in the coronary blood during vagal stimulation. One would therefore expect acetylcholine to reduce the coronary blood flow. On the contrary, this chemical dilates the coronaries in mammals, an effect which is not abolished by atropine, it has a pronounced constrictor action, however, upon the coronaries of birds (see also Gregg).

The coronary dilatation and consequent increase in flow which follow sympathetic stimulation or adrenaline administration are what one should be led to expect from the known emergency function of the sympatho-adrenal system. Through the action of this mechanism the efficiency of the body in the performance of work or in combating the dangers of the environment

is raised to a high level. The mean arterial blood pressure is elevated as a result of vasoconstriction in the skin and splanchnic areas, while the vessels of the contracting skeletal muscles and brain are dilated, the force of the cardiac contraction is augmented. It would be odd indeed if such effects, directed toward the general improvement of the circulation during muscular effort, were antagonized by a coincident constriction of the coronary arteries. But perhaps the question of the sympathetic action upon the coronaries is really of no great practical importance since the response of the coronaries in dogs after sympathetic denervation is just as great as in normal animals, nor is the dilator effect following stimulation of the stellate ganglion pronounced.

The coronary vessels are also supplied with *afferent* fibers from both the vagus and the sympathetic.

Coronary reflexes initiated from the abdominal and thoracic viscera have been demonstrated. Thus, stimulation of the abdominal organs in animals has been shown to cause coronary constriction. A coronary dilator reflex is initiated by a rise of pressure in the right auricle or vena cava (see coronary reflex below).

(3) **ALTERATIONS IN THE OUTPUT OF THE HEART**
Increase in the venous inflow, and so in the minute volume of the heart, the mean aortic pressure being kept constant, markedly augments the flow through the coronary system if the vagus nerves are intact. On the contrary, a rise in the cardiac output with a constant aortic pressure, exerts a negligible influence upon the coronary circulation after the vagus nerves have been cut. The calibers of the coronary vessels are evidently widened as a result of a reflex inhibition of the constrictor tone of the vagus. This response, which is known as the *coronary reflex*, is not abolished by excision of the stellate ganglia and the section of other sympathetic cardiac fibers. Both limbs of the reflex would appear, therefore, to be in the vagus.

(4) **OXYGEN LACK AND CARBON DIOXIDE EXCESS**
Anoxia increases the coronary flow very greatly. Reduction in the oxygen saturation of the arterial blood below 20 per cent was shown by Hilton and Eichholtz to cause maximal dilatation of the coronary vessels and a five-fold increase in flow (fig. 28.11). The addition of cyanide to the coronary blood exerted a similar effect. Carbon dioxide and lactic acid in the absence of a reduced

oxygen supply caused only a very moderate dilatation of the coronary vessels. Reactive hyperemia (p. 318) is a very pronounced phenomenon of the coronary circulation (Katz and Lindner).

(5) **VARIATIONS IN HEART RATE** It is to be expected that, since the coronary arteries fill during diastole and are compressed during systole, an increase in heart rate, which causes shortening of diastole relatively to systole, will decrease the coronary flow. In the heart lung preparation a diminution in coronary flow does occur when the rate of beating is so great that the total time occupied each minute by the diastolic periods (minute-diastole) is much reduced. When the rate is very slow, other factors remaining constant, the flow is augmented. Within the physiological range, however, a change in heart rate (70 to 180) appears, in itself, to have little effect upon the coronary flow. In an animal in which the innervation of the heart is intact, there are indications that the coronary flow is actually increased during cardiac acceleration (see Essex).

(6) **DRUGS** Nitrites, cyanides, caffeine, camphor, adenosine and mecholyl increase the coronary flow. The action of histamine varies in different species, the drug exerting a constrictor action upon the coronaries in the rabbit but a dilator effect upon those of the cat. By means of the thermostromuhr, Essex and his colleagues have investigated the action of a number of drugs on the coronary flow in unanesthetized dogs. Histamine, niketamide (pyridine- β -carboxy diethylamide), atropine and nembutal were found to increase the flow from 60 to 100 per cent, pituitrin decreased it by as much as 80 per cent (see ch. 57). Thyroxine increased the flow by 250 per cent. The effects of adrenaline and acetylcholine upon the coronary blood flow have been mentioned. *Khellin*, the active principle of a plant (*Ammi visnaga*) indigenous in Eastern Mediterranean countries, has been reported by Anrep and his colleagues to be a powerful coronary dilator. It has no

effect upon blood pressure or pulse rate. A dose of 2 mg per kilogram of body weight was found to cause a four-fold increase in the coronary flow of dogs. It appears to act directly upon the vascular wall. Alcohol, according to Sulzer, when it reaches a concentration in the blood of 0.1 per cent or more causes constriction of the coronary vessels and a reduction of the coronary flow. Drugs which increase the coronary flow in normal hearts have apparently little or no effect in increasing the blood supply to an infarcted area.

DISEASE OF THE CORONARY ARTERIES

The coronary arteries are very frequently the site of degenerative changes—atheroma and sclerosis—especially after the fifth decade of life. These changes, when advanced, lead to narrowing of the arterial lumina and a gradual insufficiency of the blood supply to the heart muscle. Milder grades of coronary disease are usually unaccompanied by symptoms or any clinical signs. Two conditions, namely, *angina pectoris* and *acute coronary occlusion*, are commonly associated with the more severe grades of coronary sclerosis. Auriculo-ventricular, or bundle branch block, pulsus alternans (chapter 24) and low voltage electrocardiograms are other consequences of coronary disease.

Angina pectoris—angina of effort

These terms are given to paroxysms of severe and often agonizing cardiac pain which occur usually as a result of exertion, and last for a few seconds or minutes. The pain is felt most usually beneath the upper or middle third of the sternum, it is frequently referred to the neck, left shoulder or arm. During the attack, electrocardiographic changes frequently appear similar to those seen in coronary thrombosis (p. 330).

Anginal attacks are most commonly precipitated by effort or some form of emotional excitement, under which circumstances the minute volume of the heart is increased but the blood supply to the myocardium is inadequate for the extra work demanded of it. An increase in the cardiac output has been demonstrated in anginal attacks. The increased work of the heart during digestion also in all likelihood accounts for the common occurrence of angina following a heavy meal. Effort at this time is particularly likely to precipitate an attack in a susceptible subject.

Several views have been expressed concerning the cause of anginal pain, e.g., spasm of the cardiac muscle, sudden stretching of the wall of

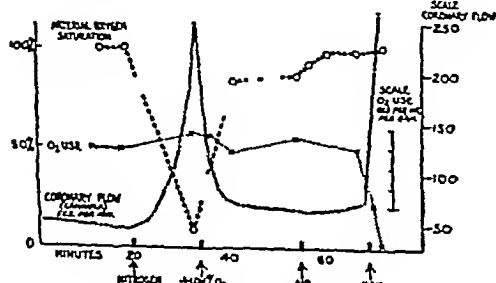


FIG. 28.11. Curves showing the relation of the coronary blood flow and of the oxygen consumption by the heart to the oxygen saturation of the blood passing through the coronary vessels (Hilton and Eichholtz).

the aortic arch or contraction of the coronary arteries. There is, however, no change in the rhythm of the heart which would indicate the existence of spasm, distension of the aorta in animals by means of a dilator introduced through the carotid does not give rise to pain, and since the coronary vessels are so often hard and stiff, spasmodic constriction of their lumina seems out of the question. The pain is evidently directly related to anoxia of the heart muscle—the oxygen supply being inadequate for the work which the heart is called upon to do. The anoxia (hypoxia) is in the great majority of cases due to narrowing of the coronary arteries—coronary insufficiency—as a result of atherosclerosis.

But, angina may occur though the coronary vessels are normal if the oxygen supply to the tissues is deficient, as in anemia or in rarefied atmospheres, and in aortic regurgitation the coronary circulation may be so diminished as a result of the low diastolic pressure (p 326) that anginal attacks occur though the heart vessels themselves are fairly healthy. Also in hyperthyroidism, when the burden thrown upon the heart is already greatly in excess of the normal, the extra cardiac work occasioned by exertion may so tax a normal coronary system to furnish an adequate blood supply to the heart muscle, that cardiac pain results. In severe anemia, the low oxygen-carrying capacity of the blood is compensated during rest by an increase in the circulation rate, the coronary flow is then adequate, the heart is able to perform the extra work entailed and the tissues do not suffer from oxygen lack (p 419). Upon exertion, however, the heart is called upon to increase its output further in order to furnish the required amount of oxygen to the contracting muscles. But the urgent demands of the heart muscle itself for extra oxygen cannot be met, since the coronary flow is already near its maximum—anoxia of the myocardium and of the tissues in general results. In cardiac hypertrophy and dilatation, anginal attacks may also occur, though the coronaries may not be diseased, and is then apparently the result of the small coronary flow *relatively* to the size of the heart (p 321). When congestive failure supervenes the greater oxygen requirement of the inefficient myocardium (p 255) is an additional factor.

Further support for the belief that anoxia is the prime factor in precipitating an anginal seizure is afforded by the observations of Katz and his col-

leagues. General anoxia induced in subjects of the disease by having them breathe an oxygen-poor atmosphere was accompanied by the appearance of characteristic electrocardiographic features (reduced amplitude or inversion of the T wave and depression of the S-T segment), a third of the subjects experienced typical anginal pain. Anoxia produced in the same way in normal subjects caused similar features to appear in the electrocardiogram, though pain was not experienced.

The pain of angina pectoris is believed to be produced in a manner essentially the same as that in which the pain of intermittent claudication is produced, namely, through the stimulation of afferent nerve endings in the myocardium by metabolic products accumulated as a result of oxygen deficiency (p 300).

The anginal attack is treated by means of nitrites (amyl nitrite, nitroglycerine, sodium nitrite) which by causing peripheral vasodilatation reduce the work of the heart and through their dilator effect upon the coronary vessels tend to increase the blood supply to the cardiac muscle. Amyl nitrite, which is administered by inhalation, acts almost instantaneously but its action is evanescent, it is therefore employed to arrest an attack. The other nitrites act more slowly but their effects are more lasting, they are used in the intervals between attacks. Adenosine compounds (p 296) are also sometimes employed for their vasodilator effects. Complete thyroidectomy, which by lowering the metabolic rate reduces the cardiac work, has been advocated in certain selected cases (p 265).

The impulses giving rise to cardiac pain pass from the heart to the central nervous system mainly via the inferior cardiac nerve, the upper four or five thoracic ganglia and the corresponding white rami and posterior nerve roots (see referred pain, p 597). Pain is not, apparently, transmitted through the vagus. Sutton and Lueth found that in conscious dogs, traction upon a ligature passed loosely around a coronary artery and brought out through the thoracic wall was followed by nausea, vomiting and evidence of pain. The response was abolished after removal of the stellate ganglia but persisted after section of the vagi.

Removal of the left stellate ganglion or the injection of alcohol around the upper five left thoracic ganglia (paravertebral injection) is sometimes resorted to for the relief of anginal pain. The afferent nervous pathways are thus interrupted and the patient is relieved of his attacks. No change in the underlying condition responsible

for the pain is brought about by these measures, and since pain is a signal which warns the patient against heart strain, its abolition is not an unmixed blessing

*Acute coronary occlusion Coronary thrombosis
Cardiac infarction*

Smith in 1918 studied the effects of ligation of various branches of the coronary vessels in dogs. Of 11 animals in which the anterior descending branch of the left coronary was ligated, 10 survived. The left circumflex artery was ligated in 14 animals, 6 survived. Ligation of the right coronary close to its origin was performed in 8 animals, only one of which recovered from the operation. In 18 animals the anterior descending branch of the left coronary and one or more branches of the left circumflex artery were tied, only 4 animals recovered from this operation, 9 died within 24 hours, 3 others died from infection and 2 from failing heart on the 15th and 17th days respectively.

Clinically, thrombosis is the commonest cause of acute coronary occlusion, though occasionally a vessel is blocked by an embolus. Thrombosis is nearly always preceded by atheromatous changes in the vascular wall. Obstruction of a main branch of a coronary artery is accompanied usually by severe pain, similar in character to that described under angina pectoris, dyspnea, nausea and vomiting, and often the signs of profound collapse. Other features of the attack are low blood pressure, fever, and leukocytosis. A fatal termination (due probably in many instances to ventricular fibrillation) either during the attack or a short time later is common. The sudden occlusion of a coronary vessel acts as such an intense stimulus to the myocardium as to induce fibrillation of the ventricles. This may be the immediate cause of death rather than failure in function of the ischemic area itself. Wiggers, however, has shown that ischemia increases the susceptibility of the myocardium to fibrillate when stimulated electrically, i.e., myocardial ischemia lowers the fibrillation threshold.

The nausea and vomiting of acute coronary occlusion, together with the fact that the attack not infrequently followed a heavy meal or a drinking bout has led in the past to many cases of coronary thrombosis being diagnosed as "acute indigestion". A proportion of subjects have suffered previously from anginal attacks. The descending branch of the left coronary is most frequently

thrombosed and for this reason has been called *the artery of sudden death*.

If death does not result from the immediate effects of the attack the area of myocardium supplied by the occluded vessel softens, undergoes necrosis and is finally replaced by a scar, or the heart may rupture through the infarcted area, death then resulting from hemorrhage. Cardiac hypertrophy may be a delayed effect of coronary occlusion (p. 257). Minor coronary branches may become obstructed without giving rise to any symptoms, the myocardial area involved being nourished through collateral vessels, or, should the occlusion of even a large branch occur gradually, collateral channels become established which are sufficient to maintain the vitality of the area of distribution of the obliterated vessel (p. 321).

The electrocardiogram immediately following an attack of coronary thrombosis often shows characteristic features which were first pointed out by Pardee. Similar features were described by Smith following coronary ligation in dogs. In occlusion of the anterior descending branch of the left coronary, the most usual electrocardiographic changes are as follows. The descending limb of the R deflection in lead I does not reach the base line but gives rise in its lower third to the T wave, i.e., the R-T interval lies well above the base line (see fig. 28 12a). In lead III, on the other hand, the descending limb of R passes below the base line but does not rise sharply again before giving rise to the T wave, the S-T interval is therefore depressed. In occlusion of the right coronary it is usual to find these characteristic features in the two leads reversed, the S-T interval being depressed in lead I and the R-T interval elevated in lead III (fig. 28 12b). The electrocardiogram changes in character as recovery progresses, gradually returning to normal though an inverted T wave in lead I (left coronary occlusion) or in lead III (right coronary occlusion) often persists. The electrocardiogram resulting from left coronary occlusion is therefore frequently referred to as the T_1 type of record (fig. 28 12c), that associated with occlusion of the right artery as the T_2 type (Parkinson and Bedford).

Probably the most characteristic electrocardiographic feature of coronary occlusion is the variability of the tracing from day to day during the period following the attack.

LOW VOLTAGE ELECTROCARDIOGRAMS These are

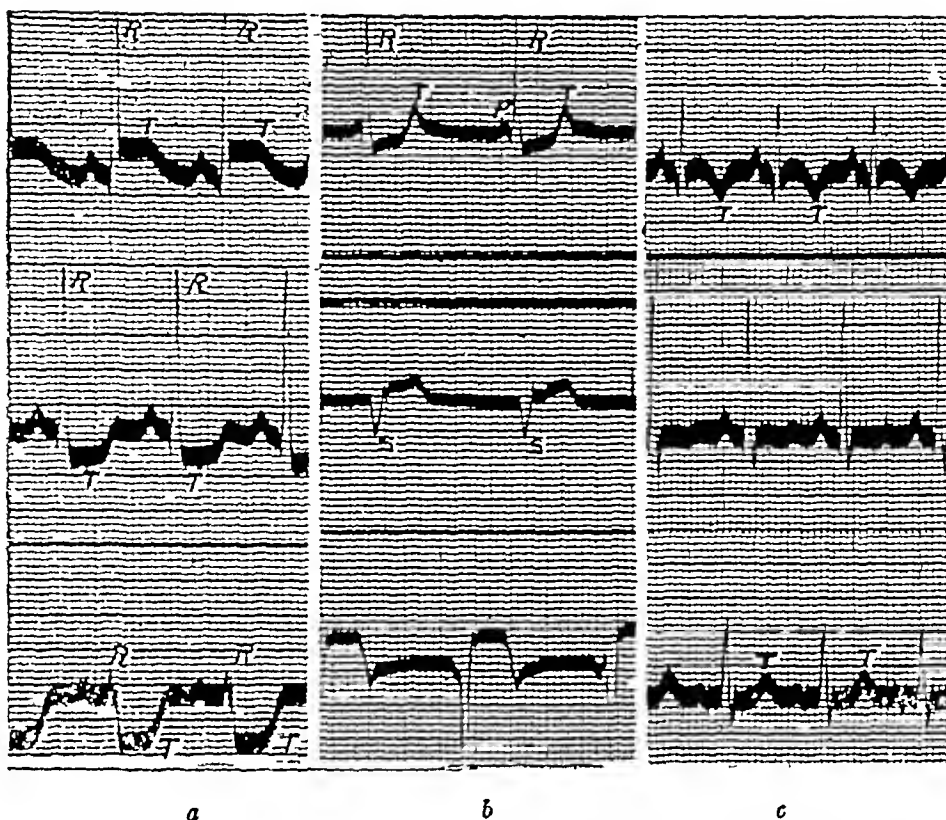


FIG 28 12 Electrocardiogram in acute coronary occlusion *a*, T₁ type of tracing, *b*, T₂ type, *c*, same patient as (*a*) 16 days later (Kindness of Dr John Hepburn)

seen most frequently in subjects of myocardial degeneration resulting from coronary disease (fig 28 13) The range of voltages for the normal electrocardiogram is given on page 220 Low voltages are also seen in a certain proportion of cases of hypothyroidism and in massive pericardial effusions Electrocardiograms showing these low voltages, except in the last two mentioned instances, are of grave prognostic significance In 68 per cent of a series of cases observed by Hepburn and Jamieson the average duration of life after this electrocardiographic feature had been first observed was 6 months

THE PULMONARY (LESSER) CIRCULATION

The blood is conveyed from the right to the left side of the heart through the pulmonary circuit and the quantity passing per minute obviously must equal the quantity flowing in the same time through the rest of the body, namely, from 4 to 6 liters under basal conditions A portion of the blood delivered to the lungs by the bronchial arteries drains directly into the pulmonary veins (see p 344) The pulmonary vessels like those of the systemic system consist of arteries, arterioles, capillaries and veins, the pulmonary arteries break up abruptly into short, wide branches The

minute vessels of the lung, unlike those of any other vascular area, are surrounded almost completely by air Though each capillary is only from a half to one millimeter in length and of the order of 10 microns in diameter, the total capillary surface exposed to the lung air has been estimated at about 140 square meters (Hufner) The vessels of the lung are highly distensible and their capacity alters with (a) changes in intrathoracic pressure incident to the respiratory movements, and (b) alterations in the minute volume of the right ventricle in relation to the resistance on the left side of the heart

The velocity of flow through the individual pulmonary vessels varies directly with the output of the right ventricle and inversely with the capacity of the vascular bed That is, the capacity of the vessels remaining constant, an increase in the output of the right ventricle increases the velocity of flow in the minute vessels (p 178) and vice versa An increase or decrease in the capacity of the vessels and thus in the total sectional area of the vascular bed, the right ventricular output remaining unchanged, slows or quickens respectively the blood flow through the pulmonary circuit

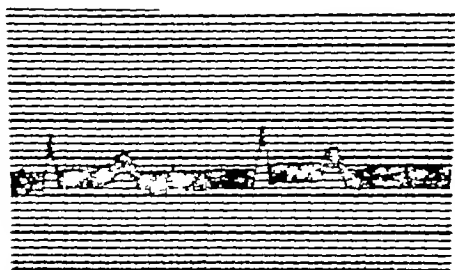


FIG. 25-15. Low voltage electrocardiogram, Type III. (Courtesy of Dr. John H. H. H. H.)

The mean arterial pressure in the pulmonary circuit amounts to only one sixth or less of that in the aorta—about 20 mm. Hg in the dog and from 12 to 30 mm. Hg in man, as determined by intracardiac catheterization; the diastolic pressure is less than one third of about 10 mm. Hg. It is evident that pressures much higher than these would disturb the balance of hydrostatic and osmotic forces of the blood in the capillaries of the lung tissue, while the blood in other capillary areas is not opposed by an external pressure (p. 31). In the dog, the average flow in the pulmonary artery is about 0.5 liters per second. The pressure in the pulmonary artery is normally near the zero level—about -3 mm. H₂O in inspiration, to +2 mm. H₂O during expiration. The tone of the pulmonary artery is a relative problem, so that a rise in the pulmonary venous pressure is readily transmitted to the arterial side, as a result of the high degree of distensibility of the pulmonary vascular bed, due mainly to the opening up of fresh capillaries, the blood contained within the lungs shows wide variations under different conditions.

Normal variations in the quantity of blood in the lungs are accompanied by little or no change in pressure. Increase in cardiac output resulting from muscular exercise, and the much greater quantity of blood contained in the pulmonary capillary bed at this time, does not alter the pressure in the pulmonary circuit of healthy persons but causes a sharp rise in those with chronic pulmonary disease. This is an indication that in the latter the distensibility of the pulmonary vascular bed is reduced. Another, however, in a small subject increases pulmonary vascular resistance and as a consequence, causes a rise in pulmonary arterial pressure. Decreased distensibility of the lungs from whatever cause would be expected to act similarly. Ordinarily, the lungs hold about 9 per cent of the total blood volume during in-

spiration and 6 per cent during expiration, under certain circumstances the quantity of blood in the pulmonary vessels may increase to 20 per cent or more of the total blood volume. Increased resistance to the flow of blood in the pulmonary veins, as in mitral stenosis or as a result of failure of the left ventricle (p. 258), raises the pressure in the pulmonary system. An abnormally high pressure in the pulmonary circuit (pulmonary hypertension) occurs also in emphysema and left ventricular failure, but does not accompany systemic arterial hypertension. Primary pulmonary hypertension is a rare abnormality. With a rise in pulmonary pressure there is engorgement of the lung vessels, the distended capillaries encroach upon the air spaces and the vital capacity, in consequence, is reduced.

The increased filling of the right heart during muscular exercise results in a greater systolic discharge from the right ventricle into the pulmonary circuit. The left ventricle does not respond instantly to the greater inflow, but only after a beat or two, or until the pulmonary venous pressure rises sufficiently to distend the ventricular cavity and stretch the muscle fibers (p. 253). The ventricle then discharges as much blood as it receives. The rise in venous pressure has, however, resulted in distension of the pulmonary vessels and an increase in the total quantity of blood contained in the lungs. But, as mentioned above, the arterial pulmonary pressure does not rise, in a healthy subject, owing to the ready distensibility of the vascular bed of the lungs.

The blood flow from the venae cavae into the heart remaining constant, a rise in aortic pressure is attended by a rise in pulmonary arterial pressure. This is not a "back-pressure" effect due to the left ventricle failing to discharge its contents adequately against the raised systemic resistance, but is the result, as shown by Anrep and Bulatao, of the greater quantity of blood returned to the right ventricle through the coronary system and, in consequence, of the larger quantity discharged into the pulmonary circuit.

THE EFFECTS OF THE RESPIRATION UPON THE PULMONARY AND SYSTEMIC BLOOD PRESSURES

The pulmonary arterial pressure falls during ordinary inspiration and rises during expiration. One should expect that, as a result of the increased flow of blood into the right ventricle during inspiration, and the greater systolic dis-

charge, the pulmonary pressure would rise during this phase of respiration. Due, however, to the traction exerted upon the circumference of the pulmonary vessels by the surrounding lung tissue, their capacity is increased. This more than compensates for the greater amount of blood entering the pulmonary circuit during the inspiratory phase. During expiration these effects are reversed. The right systolic discharge is less but the capacity of the vascular bed of the lungs is at the same time reduced, an upward swing in pulmonary arterial pressure occurs. With *maximal* expansion of the lungs or during a forced expiration with the glottis closed (Valsalva's experiment, p. 169) the vessels are strongly compressed by the surrounding lung tissue and the pulmonary arterial pressure rises sharply.

The increased capacity of the pulmonary vessels during inspiration reduces, momentarily, the flow of blood into the left auricle, the consequent reduction in the systolic discharge of the left ventricle causes a fall in aortic pressure. After a few beats of the right ventricle the greater capacity of the pulmonary vessels again becomes filled and the flow of blood into the left chambers of the heart increases, the aortic pressure rises. The succeeding expiration, by reducing the capacity of the pulmonary vessels, drives blood to the left side and further increases the discharge into the aorta, the systemic pressure, in consequence, continues its rise until near the end of the expiratory phase. The large undulations which appear in the blood pressure tracings of animals are due to these effects. If the respiratory movements and the systemic blood pressure are recorded simultaneously it is found that the blood pressure commences to fall at the commencement of inspiration and reaches its lowest point in the latter half of this phase, the blood pressure tracing then commences to rise and reaches its maximum toward the latter part of expiration⁵ (fig. 28 14).

NERVOUS REGULATION OF THE PULMONARY VESSELS

The pulmonary vessels receive vasoconstrictor fibers through the *sympathetic*. The first definite evidence of a vasoconstrictor supply to the vessels of the lung was secured by Bradford and Dean who stimulated the peripheral ends of the thoracic nerves from T2 to T7 in the dog and observed a rise in pulmonary arterial pressure without any change in heart rate or in systemic blood pressure.

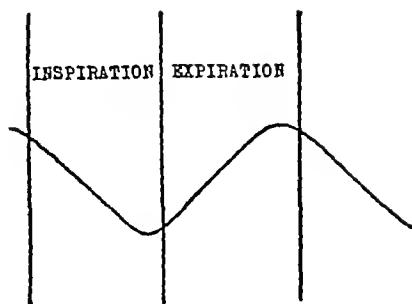


FIG 28 14 Curve of arterial blood pressure (systemic), showing the effect upon it of the respirations. For the sake of simplicity the smaller waves due to the heart beat, and normally seen superimposed upon the respiratory waves, have been omitted.

Daly and Euler have obtained evidence that the sympathetic of the dog also contains vasodilator fibers. Adrenaline causes constriction of the pulmonary vessels, except in high dilution when it causes vasodilatation.

The *vagus* of the dog carries both vasoconstrictor and vasodilator impulses to the lung. The distribution of the vasoconstrictor and vasodilator fibers to the pulmonary vessels varies considerably, however, in different species. Though in the dog, as just stated, both the sympathetic and the *vagus* contain vasoconstrictor as well as vasodilator fibers, sympathetic stimulation gives predominantly vasoconstrictor⁶ and vagal excitation vasodilator effects. In the rabbit, on the other hand, only constriction of the pulmonary vessels can be obtained by stimulation of the *vagus*, this effect is abolished by atropine and enhanced by eserine. Further evidence of the vasodilator action of pulmonary parasympathetic fibers in the dog and of their vasoconstrictor action in the rabbit is provided by the action of acetylcholine, this parasympathomimetic drug causes dilatation of the pulmonary vessels of the dog, and constriction of those of the rabbit.

It is probable that the innervation of the pulmonary vessels is of little physiological importance. Indeed, the need of vasomotor control of the pulmonary circulation is not evident.

PULMONARY CARDIOVASCULAR REFLEXES

Evidence for the existence of proprioceptors in the vascular bed of the lungs has been furnished by the

⁶ One is tempted to doubt that these experimental results reflect what actually occurs in normal unanesthetized animals, for the sympathetic, which is called into play in preparing the body in emergencies, would then act to impede the blood flow through the lungs. Furthermore, in muscular exercise a discharge of impulses over sympathetic paths occurs, yet the pulmonary arterial pressure does not rise as it would be expected to do were pulmonary vessels constricted.

⁵ See Heinbecker.

experiments of Schwiegk and of Daly and associates. The latter perfused the pulmonary and the systemic vessels separately with oxygenated blood. Increased pressure in the pulmonary circuit caused a slight fall in systemic arterial pressure and an increase or a decrease in the heart rate. The fall in systemic blood pressure was more readily produced by raising the pressure in the pulmonary veins, as may follow obstruction to the outflow from the left auricle. The effect is evidently purely reflex in character, for it is abolished by section of the vagosympathetic nerves. A reflex of this character may play some part in the fall of systemic blood pressure which occurs in left ventricular failure and in pulmonary embolism.

Inflation and deflation of the lungs within the physiological range causes, respectively, acceleration and slowing of the heart. This respiratory arrhythmia is often very pronounced in young persons (p. 240). The mode of its production has been studied experimentally by a number of investigators. According to Anrep, two mechanisms are concerned, reflex and central. The reflex effect is through afferent pulmonary terminals, the vagus center and efferent cardiac vagus fibers. The impulses set up by inflation cause depression of the tone of the vagus (cardio-inhibitory) center with consequent increase in heart rate. The reflex effect is abolished by sectioning either the pulmonary (afferent) or the cardiac (efferent) vagal fibers. The afferent impulses exert no influence upon the cardio-accelerator mechanism, for excision of the stellate and upper thoracic ganglia does not alter the response. The reflex is apparently initiated through the stretch receptors in the visceral pleura or in the layer of lung tissue immediately subadjacent to it. These receptors adapt rapidly and if the inflation is protracted, secondary slowing of the heart occurs. Extreme inflation of the lungs causes cardiac slowing rather than acceleration. The slowing of the heart caused by deflation of the lungs is accompanied by depression of auriculo-ventricular conduction, in a heart in which there is already some delay in conduction over the bundle, deflation may cause complete heart block. The central factor in the mechanism leading to cardiac acceleration following inflation of the lungs is generally believed to be due to the radiation of impulses from the respiratory to the cardio-inhibitory center. It is not abolished by section of the pulmonary fibers of the vagus or paralysis of the respiratory movements by means of curare.

THE CIRCULATION THROUGH THE LIVER

The liver receives blood from two sources—from the gastro-intestinal tract, spleen and gall-bladder through the *portal vein*, and from the aorta through the *hepatic artery*.

The portal vein differs from other veins in that it

divides into numerous branches which ultimately form a rich capillary network within the liver substance. In this it resembles an artery, but unlike an artery it is interposed between two capillary beds—one, as just mentioned, in the liver, the other in the splanchnic area. It drains the capillary beds of the gastrointestinal tract, spleen and pancreas, but not that of the kidney. The primary branches of the portal vein upon entering the liver divide into vessels which run between the hepatic lobules (fig. 28-15). These—the *interlobular veins*—give rise to capillary like vessels called the *hepatic sinusoids*. On its either side each sinusoid is separated from a bile capillary by a single layer of hepatic cells (liver cords) with which the blood comes into direct contact, an endothelial wall being absent. Blood plasma penetrates into the liver cells through fine intracellular canaliculi. The Kupffer cells (p. 105) are found in the walls of the sinusoids. The sinusoids converge toward the center of the lobule—like the spokes of a wheel toward its hub—where they empty into a wider channel running perpendicularly to them, and called the *central vein*. Each sinusoid has a sphincter at its inlet, and another at its junction with the central vein. The walls of the venules of the portal vein contain little or no smooth muscle, and are, therefore, incapable of active changes in caliber. The hepatic veins, on the contrary, are well supplied in this respect (see below). The central veins of neighboring lobules join in groups to form *sublobular veins* which unite in turn to form *hepatic veins*. After a series of unions of the latter, larger vessels (still called *hepatic veins*), are formed which empty into the inferior vena cava. The hepatic veins opening into the inferior cava are from 5 to 13 in number, one or two only drain the left hepatic lobe, the remainder carry blood from the right lobe.

The hepatic artery is distributed mainly to the fibrous tissue of the capsule and interlobular septa, but also to the hepatic parenchyma through the mixing of its blood with the portal blood, for the hepatic artery has three communications with the portal system, (a) its terminal twigs open into the sinusoids, (b) into the portal vessels just proximal to the sinusoids, and (c) arterio-venous channels enter the portal system some distance proximal to the sinusoids. The blood conveyed to the liver through the hepatic artery constitutes about 20 per cent of the organ's total blood supply, and like arterial blood elsewhere has a high oxygen content. The blood of the portal system is only about 50 per cent saturated with oxygen. The oxygen consumption per gram of hepatic tissue varies between 0.024 and 0.065 cc per minute. Though the hepatic artery supplies only about 20 per cent of the blood to the liver it is responsible, owing to its relatively high oxygen tension, for delivering the greater part of the oxygen consumed by the organ. Necrosis of the liver follows shortly after ligation of the hepatic artery. Markowitz and his associates have found, however, that necrosis of the liver, following

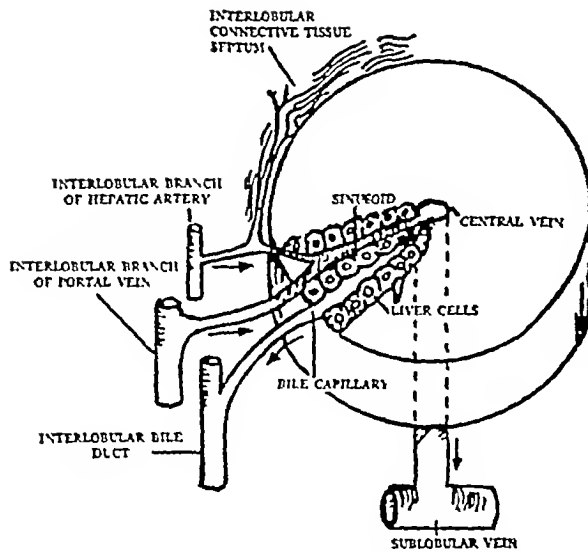


FIG 28 15 Diagram of the hepatic circulation

ligation of the hepatic artery, is due to infection with anaerobic bacteria, and that the supply of arterial blood may be entirely cut off with impunity if the animal is treated with penicillin. They suggest that the chief function served by the hepatic artery is the maintenance of a sufficiently high oxygen tension in the hepatic circulation to prevent the growth of such microorganisms.

The circulation is intermittent over large territories of the liver, in about 75 per cent of the liver substance the circulation, under ordinary circumstances, is inactive, its sinusoids being packed with red cells and apparently serving merely as storage depots. As might be expected, many more sinusoids show circulatory activity during digestion, and the blood flow through the liver increases greatly.

The blood pressure in the portal vein, as measured in experimental animals, is from 8 to 12 mm Hg. In the dog and probably in man the muscular coat of the hepatic veins near their entrance into the vena cava is highly developed. When contracted the smooth muscle in this situation, which Dale and his associates refer to as the "hepatic sphincter", may cause a marked increase in the resistance to the hepatic blood flow. Through this arrangement the hepatic veins serve the purpose of an adjustable sluice, by means of which the quantity of blood in the liver is varied considerably under different conditions. The liver thus functions as a reservoir within which blood at one time may be pooled, or from which at another time an extra quota of blood may be delivered into the general circulation. Barcroft, Nisimaru and Ray point out, however, that the liver is not a storehouse in the sense that the blood

is in a cul de sac and out of circulation, as in the case of the spleen (p 70). The time taken for a red cell to traverse the hepatic channels (6 to 10 seconds) is only a fraction of the time that it spends in passing through the spleen or even the skin. It was found, for example, that 30 minutes after an animal had breathed carbon monoxide, samples of blood taken from the femoral artery, portal vein, hepatic veins and from the scratched surface of the liver, all had practically the same CO saturation. That is, there had been no greater retention of CO by the hepatic blood, and consequently no significant slowing of the flow through the liver.

The "hepatic sphincter" is constricted by peptone and histamine, and relaxed by adrenaline or sympathetic nerve impulses. The great engorgement of the liver and splanchnic area in dogs following histamine poisoning (histamine shock) is illustrative of the effect of this drug upon the hepatic veins. The splanchnic congestion seen in anaphylactic or peptone shock in this species, is also due, apparently, to constriction of the hepatic veins and the damming back of blood in the portal system. Adrenaline, on the other hand, according to the observations of Rein and his associates, causes the immediate delivery from the liver into the general circulation, of a weight of blood equal to from 26 to 59 per cent of the weight of the liver itself. Adrenaline, at the same time as it relaxes the sphincters of the hepatic veins, appears to constrict the finer tributaries within the liver, and causes, as well, the liberation of fluid from the hepatic cells.

Acetylcholine appears to have little effect upon the hepatic blood flow, but the cholinergic drug β -methylcholine stimulates the muscle of the hepatic veins, and, after a brief increase in venous outflow, there is a rather prolonged period of reduced flow. Stimulation of the vagus nerve causes the appearance of acetylcholine in the effluent blood, whereas an adrenaline-like substance is found after stimulation of the hepatic plexus (sympathetic). Infusions of hypertonic solutions of glucose or of saline increase the hepatic blood flow.

The blood flow through the liver is also varied through nervous influences acting upon the intra-hepatic vessels and upon the vessels of the splanchnic area. Variations in the caliber of the former will alter the resistance to the flow through the liver. Variations in diameter of the splanchnic

vessels, on the other hand, increase or diminish the quantity of blood reaching the liver through the portal tributaries

The blood flow through the liver in man can be determined upon the direct Fick principle (p 267) by obtaining samples of arterial systemic blood and of hepatic venous blood (by catheterization of an hepatic vein) The samples are analyzed for their content of urea (produced in liver), and the total urea excretion in the urine per minute determined From these data the hepatic blood flow is calculated Bradley and his associates employ bromsulphthalein, which is removed selectively by the liver, the dye is continuously transfused, and its amount in the arterial and in hepatic vein blood ascertained Thus, if 50 mg of the dye are infused per minute and the difference between the amounts in arterial and hepatic vein blood is, say, 0.50 mg per 100 cc., then the hepatic blood flow is $(50/0.50 \times 100 =)$ 1000 cc per minute The hepatic blood flow in man is around 1500 cc per minute.

OBSTRUCTION OF THE PORTAL CIRCULATION

In cirrhosis of the liver (portal cirrhosis), the compression of the intrahepatic vessels by fibrous tissue raises the resistance to the flow of blood through the liver The hepatic vascular bed is much reduced As a result of the rise in pressure in the portal system, the blood is forced into channels, inconspicuous under normal conditions, connecting the portal vein and the systemic veins The chief among these by-passes for the return of blood from the splanchnic area to the right auricle are provided by anastomoses between (a) the coronary vein of the stomach and the esophageal veins, (b) the superior hemorrhoidal branches of the inferior mesenteric vein and the middle and inferior branches of the internal iliac (hypogastric) veins, (c) the epigastric veins (superior and inferior) and the portal vein through the paraumbilical veins of Sappey When these communications are fully established the epigastric veins appear as a group of large tortuous vessels of the abdominal wall in the region of the umbilicus (*caput medusae*), (d) the radicals of the portal vein in the intestines and the inferior vena cava, through the retroperitoneal veins of Retzius

Portal hypertension is also seen in congestive splenomegaly (Banti's syndrome, p 72) Surgical treatment consists in anastomosing the portal vein to the inferior vena cava

ECK FISTULA This is the communication produced for experimental purposes between the portal vein and the inferior vena cava, and the exclusion of the liver from the portal circulation The two veins are anastomosed side by side and the portal vein then ligated between the anastomosis and the liver, the blood from the portal area is thus turned directly into the vena cava The hepatic artery is, of course, left intact This operation has been employed in certain studies where it has been desired to exclude hepatic function Such an object is not attained, however, for the only satisfactory means of abolishing hepatic function is to remove the liver

EXTIRPATION OF THE LIVER Ordinarily, death occurs within a very short time after removal of the liver The rapid death is not due to the loss of the essential hepatic functions, but simply to the fact that since the portal vein must be tied, the return of blood from the splanchnic area to the systemic circulation is blocked The vessels of the intestines and spleen become intensely engorged, the animal dying as a result of the great reduction in its circulating blood volume, or, as the expression goes, it bleeds into its own splanchnic vessels

Mann and Magath have devised a three-stage operation which makes it possible to remove the liver completely from dogs and have them survive for periods varying from 20 to 36 hours The first stage of the operation consists of anastomosing the portal vein and the inferior vena cava and ligating the latter vessel above the anastomosis, but below the entrance of the hepatic veins All the blood from the lower part of the body is thus directed through the liver This *reverse Eck fistula*, as it is called, and the consequent great increase in the blood flow to the liver, elevates the portal venous pressure Anastomotic channels gradually open up between the portal and systemic systems through which a large part of the blood is diverted After a few weeks when the collateral circulation has become well established, the portal vein is tied The final operation, performed after the lapse of another two weeks, consists of tying the hepatic artery and removing the liver, which has now shrunk considerably in size

Marlowitz and Soskin have developed a simplified technique for the removal of the liver which does not involve the anastomosis of the portal vein to the vena cava They apply ligatures to these vessels which constrict, but do not com-

pletely occlude them. The increased resistance thereby induced results in the development of the collateral channels. After these have become established the portal vein and inferior vena cava are completely occluded, and the liver removed in the usual way.

It is also possible to extirpate the liver at one sitting (Markowitz and colleagues) by connecting the infra- and supra-hepatic sections of the vena cava by means of a cannula of pyrex glass (which, as shown by Firor, does not cause blood clotting). The vena cava below the liver is opened. The cannula is inserted into the vein and passed upward through its intra-hepatic portion into the supra-hepatic portion, the vein is then tied around the lower end of the cannula. The portal vein and inferior vena cava are next anastomosed. The vena cava is then tied around the upper end of the cannula. Finally, the portal vein and hepatic artery are ligated and the liver removed, the hepatic tissue being stripped away from the reconstructed vena cava.

THE CEREBRAL CIRCULATION

ANATOMICAL DESCRIPTION

Blood enters the cranium through the *internal carotid* and *vertebral arteries*. The brain of man receives the greater part of its blood supply through the internal carotids. The two vertebrals unite to form the *basilar artery* which runs forward in the median groove on the under surface of the pons and divides into the two *posterior cerebral arteries*. Each internal carotid divides near the lateral border of the optic chiasma into the *anterior* and *middle cerebral arteries*. Communicating vessels unite the posterior and middle cerebrals on either side and the two anterior cerebrals. Thus a vascular ring—the *circle of Willis*—is formed at the base of the brain around the optic chiasma, the tuber cinereum, infundibulum and corpora mammillaria. Through the circle of Willis very free communication between the internal carotids and the vertebrals is established. Owing to the presence of these anastomotic channels both carotids can be tied in the monkey without serious interference with the blood supply to the brain. Ligation of both carotids and one vertebral in this animal results in stupor and death. The dog survives after both internal carotids and the two vertebrals near their origins have been ligated. Blood then reaches the circle of Willis through anastomoses between the spinal branches of the vertebrals and the deep cervical artery, and also, as shown by Bouckaert and Heymans, through an ophthalmic branch of the internal maxillary artery (branch of external carotid) which communicates with the internal carotid within the skull. This latter communication is capable of maintaining the viability of the brain, though all other channels through which blood could enter the

cranial cavity have been tied. Temporary occlusion of all the main cerebral arteries in man has been described by Kabat and associates. The main effects are fixation of the eyes in the mid-line, tingling sensations, constriction of the visual fields and loss of consciousness in about 7 seconds. After restoration of the circulation, mild tonic and clonic seizures occur. After occlusion for 100 seconds and restoration of the circulation, there is rapid return to consciousness, no lasting effect was observed. The corneal reflex was lost in less than 10 seconds after the circulation had been arrested, the abdominal and plantar reflexes were abolished, and the Rossolimo and Hoffmann reflexes became positive. In most instances, ligation of one common carotid in man causes no ill effects owing to the rich anastomoses between the two external carotids and between the external carotid with the internal carotid of the opposite side. Syncope, convulsions, etc. reported as resulting from compression of one common carotid were in all probability due to a carotid sinus reflex (p. 287) rather than to mechanical interference with the cerebral blood supply. Compression of both common carotid arteries in man is followed within a few seconds by unconsciousness.

The anterior and middle cerebral arteries, branches of the internal carotid, and the posterior cerebral branches of the basilar, after giving off branches which supply chiefly the basal ganglia and hypothalamus, pass to the surface of the cerebrum. Here they ramify in the pia mater and send twigs into the underlying substance of the hemispheres. The anterior cerebral courses forward from its origin to the commencement of the longitudinal fissure and then arches backward over the corpus callosum. It is distributed over the medial surface of the frontal and parietal lobes, an adjacent band of cortex on the lateral aspect of the hemisphere, and the cortex of the cingulate gyrus. The middle cerebral artery reaches the surface of the hemisphere through the Sylvian fissure. Among its basal branches are twigs—the *lenticulo striate*—to the lentiform nucleus and internal capsule. One of these, larger than its fellows, not infrequently ruptures and on this account is referred to as the "artery of cerebral hemorrhage". The cortical distribution of the middle cerebral includes the anterior pole of the temporal lobe, the lateral part of the orbital surface of the frontal lobe, the insula, and the large area of the lateral surface of the hemisphere not supplied by the anterior or posterior cerebrals, that is, the greater part of the middle and inferior frontal gyri, the lower two-thirds of the pre- and post-central gyri, the supramarginal and angular gyri, and the superior and middle temporal gyri. The posterior cerebral artery curves backwards between the cerebral peduncle and the uncus. It supplies the uncus and hippocampal gyrus on the medial aspect of the hemisphere, the fusiform and inferior temporal gyri, and the cortex on the medial, inferior and lateral surfaces on the occipital lobe.

The vertebral artery on each side gives off a little below the pons a *posterior inferior cerebellar* branch which supplies the medulla. It sends twigs to the glossopharyngeal, vagus and accessory nuclei, to the spino-thalamic, spino-cerebral and rubro-spinal tracts, and, probably, to the spinal root of the trigeminal nerve. The basilar artery, formed by the union of the two vertebrals gives off on each side an *anterior inferior cerebellar* branch which ramifies over the under surface of the cerebellar hemisphere, and a *superior cerebellar* branch which is distributed to the superior surfaces of the vermis and cerebellar hemisphere.

The blood is returned from the brain through the *cerebral veins* which course over the surface of the hemispheres, and receive as tributaries the small veins which drain the capillary bed in the brain substance. The cerebral veins drain into the large *dural sinuses*—*superior and inferior sagittal, cavernous, straight*, etc.—which lie enclosed between the outer and inner layers of the dura mater. The sinuses possess no valves. Their triangular cross-section and the support afforded by their dural envelopes render them relatively resistant to compression. The venous blood ultimately finds its way into the two *transverse sinuses*. The transverse sinus on either side becomes continuous at the jugular foramen with the internal *jugular vein*.

The network of veins which begins with the dural sinuses in the posterior cranial fossa forms a longitudinal plexus which extends throughout the entire length of the spinal column and communicates at each intervertebral space with veins within the thorax and abdomen. This plexus constitutes a route through which the blood can be returned from the cranial cavity to the superior and inferior venae cavae. That this alternative channel is of great functional importance in man is shown by the fact that both internal jugulars can be ligated simultaneously without serious consequences. Even complete occlusion of the superior vena cava by a thrombus may occur with relatively little ill-effect. Other less important channels through which blood may be returned from the cranium are the emissary veins of the skull which connect the dural sinus with the veins of the scalp, and communications between the veins of the orbit and the pterygoid plexus of veins.

The gray matter is much more vascular than the white matter. In the human cerebral cortex the total capillary length is about 1 meter per cu mm of tissue, and only around 200 mm per cu mm in the underlying white substance. The cerebral vessels are not "end arteries" but form a continuous, freely communicating network through which a corpuscle might conceivably travel from the frontal to the occipital lobe

by rigid walls and filled by brain substance, cerebro-spinal fluid and blood. The total volume of the cranial contents, which are incompressible, can therefore alter relatively little and to an extent which is determined simply by the degree of bulging of the membranes at the occipito-atlanto-toid joint and between the vertebrae.⁷ Nevertheless, under certain circumstances considerable reciprocal volume changes between the brain substance, cerebro-spinal fluid and blood may occur. That is, any increase or decrease in one of these is accompanied by a reverse change in one or both of the others. The intravenous injection of hypertonic (30 per cent) saline, for example, causes, through the changes in osmotic relationships, a shrinkage of the brain, a fall in cerebro-spinal fluid pressure from around 10 mm Hg to a negative value (–10 mm Hg), but a very moderate fall in the pressure of blood in the superior sagittal sinus. As a consequence, reversal of the normal relationship between cerebro-spinal fluid and venous cerebral pressures results, the former being reduced considerably below the latter. The creation in this way of a negative cerebro-spinal fluid pressure is strong evidence that the cranial contents are contained within a "closed box." Hypotonic solutions cause an increase in brain volume and a rise in cerebro spinal fluid pressure and of the pressure in the superior sagittal sinus. The normal relationship between cerebro-spinal fluid and cerebral venous pressures is therefore maintained.

The vessels of organs such as the kidney, liver or muscles, etc., can dilate widely and the whole organ expand (see fig 27 14, p 295). Consequently a greater mass of blood can be held by these structures at one time than at another, and it is possible for a greater volume of blood to flow through them without the velocity of flow through the individual vessels being increased. Since the mass of intracranial blood cannot be very much greater at one moment than another, a greater oxygen supply to the brain is assured largely by an increase in the speed of the blood through the individual vessels rather than by an enlargement in the total capacity of its vascular bed. The cerebral circulation time (carotid to jugular by the radium emanation method) is around 3 seconds.

⁷ The conception of the invariability of the volume of the intracranial contents as a result of the rigidity of the cranial boundaries, originated with Alexander Monro of Edinburgh (1783), and was elaborated by Kellie (1824), it is therefore commonly known as the Monro-Kellie doctrine.

FACTORS REGULATING THE CEREBRAL BLOOD FLOW

The adult cranial cavity and vertebral canal form a space which is almost completely enclosed

The velocity of the blood through the intracranial vessels is determined by the pressure difference (pressure gradient) between the cerebral arteries and the cerebral veins. The pressure in the former vessels is ordinarily proportional to the systemic arterial pressure (see below, p. 337 and 340). The pressure in the larger cerebral arteries is about 65 mm. Hg diastolic, and 100 mm. Hg systolic. The pressure in the capillaries is about 13 mm. Hg. The pressure in the intracranial veins though lower than, varies with the cerebro-spinal fluid pressure. The venous pressure in the recumbent position is from 6 to 8 mm. Hg, i.e., approximately the same as the pressure in the median basilic vein, but becomes reduced nearly to zero in the erect position. Changes in venous pressure at the right auricle are transmitted to the internal jugular vein and so to the intracranial veins. In consequence, the pressure in the venous sinuses varies as a result of the changes in intrathoracic pressure occurring during the respiratory cycle—increasing during expiration and decreasing during inspiration. Failure of the right ventricle, thrombosis of the transverse sinus or obstruction of the internal jugular vein, or of the superior vena cava, are among some of the pathological conditions which result in an increase in the cerebral venous pressure. A fall in arterial pressure or a rise in cerebral venous pressure will tend to slow the blood flow through the brain, reverse changes will tend to increase the cerebral blood flow. However, apart from the postural and respiratory variations just mentioned, the cerebral venous pressure remains fairly constant under ordinary conditions. The intracranial blood flow is, therefore, determined largely by the height of the general arterial blood pressure. Until recently this was considered to be the sole factor regulating the flow through the brain, the cerebral vessels themselves were thought to play a purely passive rôle.

The effect of gas tensions on the cerebral vessels. The gas tensions of the blood are potent factors in the control of the cerebral circulation. An increase in carbon dioxide tension (breathing 7 per cent CO_2) has a profound vasodilator effect upon the cerebral vessels, it is capable of causing changes in vascular calibers independently of alterations in the general blood pressure, and by dilating the vessels may increase the cerebral blood flow by 40 per cent. Low carbon dioxide tension has, in man, a vasoconstrictor effect. Anoxemia has been found by most investigators to exert a pronounced vasodilator

action, an increase in oxygen tension causes a decrease in blood flow through the brain. Acids act upon the intracranial vessels like high tensions of carbon dioxide or low tensions of oxygen and alkalis like low CO_2 tensions or high oxygen tensions.

THE NERVOUS CONTROL OF THE CEREBRAL VESSELS. Gulland over fifty years ago (1898) reported having discovered nerve filaments upon the blood vessels of the pia mater. Huber a year later described medullated and non-medullated nerves ending on these vessels. He considered the medullated filaments to be sensory, the non-medullated, vasomotor in function. Nerve fibers going to the blood vessels within the brain substance have also been described by others (Kolliker, Clarke) and more recently by Penfield.

Though Wiggers had demonstrated that the pial vessels reacted to adrenaline by constriction, it is only within the last few years that definite evidence of the nervous control of the intracranial vessels has been secured, it is now definitely established that vasoconstrictor impulses are conveyed by the sympathetic.

The pial vessels have been observed by Forbes and Wolff through a glass window screwed into a trephine hole in the skull, the space between the glass and the brain surface being filled with Ringer's solution. Upon stimulation of the sympathetic, constriction of pial and dural vessels was observed accompanied by a rise in systemic arterial pressure, section of the sympathetic was followed by dilatation of the exposed vessels. Stimulation of the central end of the cut vagus, aortic or sinus nerves or of the facial nerve at the geniculate ganglion, resulted in dilatation of the pial vessels and a fall in systemic pressure. The vessels of the pia mater were not altered in caliber by a rise in the systemic arterial blood pressure unless this exceeded 60 mm. Hg, dilatation of the vessels apparently passive in nature was then observed. These experiments show that the superficial cerebral vessels can constrict or dilate quite independently of, or indeed in spite of, a rise or fall in systemic pressure, and prove conclusively the existence of a nervous mechanism in the control of the intracranial circulation. This control is not essentially different from that governing the caliber of the systemic vessels except that the cerebral vessels, according to Schmidt, are held in a state of tonic dilatation rather than in one of tonic constriction, as prevails in the vascular system of the rest of the body. Reduction of vaso-

dilator tone is probably of greater importance than sympathetic impulses in bringing about constriction of the cerebral vessels. Though quite definite, the vasoconstrictor action is weak as compared to the response of extracranial vessels to sympathetic stimulation and to the stimulation of vasodilator fibers to the cerebral vessels. Stimulation of the cervical sympathetic causes a reduction in caliber of the latter of only 10 to 30 per cent in contrast to an 80 to 90 per cent reduction in diameter of the vessels of the skin to sympathetic stimulation.

A fall in systemic blood pressure below a certain critical level (about 60 mm Hg) causes dilatation of the cerebral vessels. This reaction apparently constitutes a safety device to maintain an adequate blood supply to the brain. The associated fall in blood pressure has been shown to be the factor responsible for the vasodilatation caused by stimulation of the vagus, aortic or sinus nerves (Forbes and associates), for vasodilatation fails to take place if the blood pressure is prevented from falling while the nerve is stimulated. On the other hand, the vasodilator response is not altered by the application of cocaine to the pial vessels, which indicates that it is not a true reflex but a reaction of the vascular walls themselves to the low intravascular pressure. Since vasodilatation following vagal excitation is simply a compensatory response to the fall in blood pressure it rarely causes any increase in cerebral blood flow. The cerebral vasodilatation caused by stimulation of the facial nerve occurs with a normal blood pressure and is abolished by cocaine of the pial vessels. It therefore appears to be a direct response of the vessels to nerve stimulation. The vasodilator fibers of the facial pass to the vessels via the geniculate ganglion, the great superficial petrosal and the internal carotid nerves. *The facial nerve is apparently the only source of intracranial vasodilator fibers.*

The observations just described are not incompatible with the conception that the quantity of blood within the cranial cavity remains approximately constant. For, changes in the calibers of the vessels in one area of the brain may coincide with changes of an opposite character in another,*

* A number of observations support the idea of regional variations in cerebral blood flow in accordance with functional demands. For example, Fulton observed in a patient that the blood flow of a vascular tumor (hemangioma) of the occipital lobe increased (as indicated by an accentuation of the bruit heard over the occipital bone) when the subject read fine

and, since the spinal column does not constitute a completely rigid encasement but contains distensible structures, small reciprocal changes in cerebro-spinal fluid may allow corresponding variations in vascular diameters. During sympathetic stimulation an average reduction in diameter of 8.5 per cent occurs and during vagal stimulation an average increase in diameter of 22 per cent. The dilatation of even a relatively large number of arterioles to this extent might cause little increase in the total quantity of blood contained at any instant within the cranium. Nevertheless, a very considerable increase in blood flow might result, for, according to Poiseuille's law (p. 140) the flow of liquid through a capillary tube (other factors remaining constant) is directly proportional to the fourth power of the diameter of the tube. Thus an increase of 22 per cent in vascular diameters would increase the blood flow 150 per cent (Cobb).

These facts throw a new light upon the physiology of the cerebral circulation. They indicate that with a constant systemic arterial blood pressure, changes in blood flow not only through one part of the brain in relation to another part, but through the brain as a whole may be brought about. Finesinger and Putnam, for example, perfused the brains of monkeys through the internal carotid with heparinized blood after tying the vertebrals and the opposite carotid. The minute volume of the inflow was reduced when the sympathetic was stimulated. It was found, however, that variations in the perfusion (carotid) pressure were more effective in varying the blood flow than was nerve stimulation with a constant pressure head. The maximum blood flow through the brain would of course result from a rise in systemic pressure accompanied by dilatation of the cerebral vessels. In this connection may be cited the observations of Heymans and Bouckaert. These observers found that a pressor reflex elicited from the carotid sinus did not involve the cerebral vessels. In this reflex the rise in blood pressure, due to

print. Also, in animals a thermocouple in the optic pathway records a rise in temperature when the eyes are illuminated, on the other hand, when the skin of the feet was stimulated, though there was no indication of increased flow in the optic pathways, a definite temperature rise was recorded from nervous tracts subserving cutaneous sensations (Gerard and Scrota) and, during muscular movements, from the motor cortex. Increased vascularity, as demonstrated by intravital injections of dye, is found in the olfactory lobes of the cat after the inhalation of a strong smelling gas.

vasoconstriction of the systemic arterioles, must obviously result in a greater flow through the cerebral vessels than would occur if these shared in the vasomotor response. In pathological hypertension however the cerebral blood flow is not greater than normal, the intracranial vessels sharing in the general vasoconstriction, and thus raising the cerebral vascular resistance. Nor is it likely that in muscular exercise the cerebral blood flow is increased.

Alterations in the blood flow through the human brain can be determined from comparisons of the oxygen (or carbon dioxide) contents of the blood entering and leaving the cranial cavity. Blood samples are taken from an artery and the jugular vein. An increase in the arterio-venous oxygen (or carbon dioxide) difference indicates a reduced blood flow, a lowered arterio-venous oxygen (or carbon dioxide) difference (i.e., jugular blood more arterial in character) indicates an increased flow. Variations in the intracranial blood flow of the human subject can also be demonstrated by inserting an electrically heated needle into the jugular vein, and measuring fluctuations in its temperature by means of thermocouples connected in series with a galvanometer. The heated element is cooled by the blood flowing around it, a rise or fall in its temperature, therefore, indicates a reduced or an increased blood flow respectively.

Ferris has attempted to measure the *total* blood flow through the human brain by an ingenious plethysmographic method (see p. 179). A wide-bore needle is introduced into the lumbar subarachnoid space. The cerebral blood flow is estimated from the displacement of cerebro-spinal fluid while the jugular veins are occluded for a brief period by means of a pressure cuff encircling the neck. Gibbs has adapted the Stewart dye dilution method to the cerebral circulation (ch. 26), the dye is injected into a carotid artery. The total cerebral blood flow in man, as measured by the former method, is 250 cc per minute during rest and 400 cc maximum, or 16 cc and 26 cc per 100 grams of brain substance. Owing to certain sources of error in this method (e.g., the escape of blood from the craniovertebral cavity through unoccluded veins) this estimate is probably much too low. Dumke and Schmidt obtained a higher figure for the monkey, namely, 60 and 110 cc per 100 grams of brain per minute. Kabat and his associates obtained an average figure of 100 cc per 100 gr per minute for man, or about one third of the car-

diac output for the entire brain (1300 to 1500 grams in weight) (see also ch. 62).

The demonstration that the cerebral vessels can vary their calibers independently of changes in systemic blood pressure also provides a physiological basis upon which temporary disturbances in cortical function may possibly be explained. It has long been suspected, for example, that transient hemiplegia, amblyopia, and possibly migraine or convulsive seizures, may be due to spasm of pial vessels and a consequent anemia of the cerebral centers. Attempts to gain information upon this question have been made by studying the effects of convulsant poisons in animals. Finesinger and Cobb observed acute constriction of the pial vessels preceding the convulsions induced by the intravenous injection of caffeine, but only vasoconstriction preceded the convulsions caused by the administration of picrotoxin or small doses of absinthe and vasodilatation resulted from the administration of a convulsive dose of camphor or of a large dose of absinthe. Nor does it appear that the convulsive action even of those drugs which cause pial vasoconstriction is directly due to reduced cerebral blood flow. Gibbs, for example, found that convulsions followed the administration of caffeine though the blood flow was increased by injecting adrenaline (see below). The general belief that cerebral anoxia causes an initial stimulation of nervous tissue with the production of convulsions, has been disputed by Schmidt who maintains that oxygen lack always depresses cerebral functions.

THE ACTIONS OF CERTAIN CHEMICALS AND DRUGS UPON THE CEREBRAL VESSELS

Histamine dilates the pial vessels and raises the cerebro spinal fluid pressure. Flushing of the cortex has been observed in man following its administration, this is probably the cause of the intense but transient headache which sometimes follows an injection of histamine. Weiss and Lennox showed also that the arterio-venous oxygen difference was reduced, i.e., the intracranial blood flow was increased, by histamine. Since the systemic blood pressure remained practically unaltered, dilatation of the cerebral vessels must have occurred. *Metrazol*, *ether*, *alcohol* and *carbon monoxide* are also dilators of the pial vessels. *Adrenaline* applied locally to the surface of the brain causes vasoconstriction. Its injection into the general circulation, however, is followed by vasodilatation. This is a passive effect due to the rise in blood pressure, that is, the latter overcomes the local effect of the hormone upon the pial vessels. The net result of adrenaline liberation under physiological conditions is therefore an increase in cerebral blood flow. *Benedrine* given by intracarotid injection reduces the blood flow to the brain. *Pitressin* has an inconstant effect but in most instances causes cerebral vasoconstriction. *Acetylcho-*

line is vasodilator but, owing to its causing a coincident fall in blood pressure, little or no increase in blood flow results, it may even cause a reduction. Most *anesthetics*, especially ether, cause vasodilatation. *Hyper tonic solutions* administered intravenously cause a brief period of vasoconstriction followed by an increased blood flow. This is apparently due to the well known action of such solutions in reducing the intracranial pressure and in causing a rise in systemic arterial pressure. *Caffeine* in large doses decreases the cerebral blood flow. *Amyl nitrite* dilates the cerebral arterioles as it does those of the rest of the body. There are very few agents known which constrict the cerebral vessels. Among them are *barium chloride* and, as already mentioned, a high tension of oxygen. A low CO_2 tension, of course, results in vasoconstriction.

THE EFFECT OF INCREASED INTRACRANIAL PRESSURE UPON THE CEREBRAL CIRCULATION

Forbes and Wolff studied in cats the effects upon the cerebral circulation of raising the intracranial pressure. They introduced Ringer's solution under pressure into the cisterna magna and observed the superficial vessels through a glass window fixed into the cranial wall. It might be thought that any considerable rise in intracranial pressure would compress the cerebral veins and arrest the circulation. No change in the vessels of the pia mater could be detected, however, until the cerebro-spinal fluid pressure had been increased to a value four or five times that of the normal. With gradual elevation of the pressure the sequence of vascular events was, —slowing of the blood flow in the veins, then dilatation of their lumina, dilatation of the arteries, slowing of the arterial blood flow, narrowing of the arteries, and

finally, complete obliteration of the vessels with consequent blanching of the cortex. It appears that as the cerebro spinal fluid pressure rises, the pressure is transmitted to the blood in the veins and through these vessels to the capillaries and arteries. Thus, though at first no change in systemic arterial pressure occurs, a higher arterial pressure is established within the cranium and the difference between arterial and venous cerebral pressures is maintained. The cerebral circulation, though slowed, is able to continue despite a high cerebro spinal fluid pressure. When the cerebro-spinal fluid pressure equals the systemic arterial pressure the circulation through the cerebral vessels, of course, becomes impossible.

The "local" rise in arterial pressure accounts for the fact that in clinical cases in which, though the intracranial pressure is considerably elevated and the systemic arterial pressure unaltered, the circulation through the brain is not cut off. It has been shown by Berens and his associates that in subjects of brain tumor a rise in the pressure in the retinal arteries may occur unaccompanied by any elevation of the systemic pressure. When the cerebro-spinal pressure continues to rise and the pressure in the cerebral arteries approaches that in the systemic vessels, the intracranial blood flow must be seriously curtailed and finally arrested unless further adjustments occur. These are brought about through the vasomotor center. The slowing of the flow through the medulla stimulates the center, a rise in systemic blood pressure results to force blood through the cerebral vessels threatened with obliteration.

The renal circulation is described in chapter 35

SECTION III. RESPIRATION

By N B T

Exclusive of parts of chapter 32

CHAPTER 29

INTRODUCTION

THE MECHANICS OF RESPIRATION

The term "respiration" refers to the gaseous interchange between an organism and its environment. The more obvious chemical features of this process are the absorption of oxygen and the elimination of carbon dioxide. All living things, excluding certain microorganisms which secure energy from dehydrogenase or similar systems, must be supplied with oxygen. The oxygen is absorbed by the blood in the lungs for delivery to the tissue cells wherein carbon is oxidized to carbon dioxide and hydrogen to water. The CO_2 is transported by the blood to the lungs and eliminated in the expired air. The exchange of the respiratory gases between the tissue cells and the internal environment, which is constituted by the fluids bathing the cells, is called *internal respiration*. This process is essentially the same as that taking place between unicellular organisms (e.g., the amoeba) or primitive multicellular forms and the aqueous environment in which they live. But internal respiration also includes dehydrogenations through the activity of various enzyme systems in the tissues. The exchange of oxygen and carbon dioxide between the blood in the pulmonary capillaries and the air in the lungs is termed *external respiration*. The study of respiration therefore involves, principally, (1) the physiological mechanisms responsible for the body obtaining an adequate supply of oxygen from the external environment, (2) the transport of oxygen from the lungs to the tissues and of carbon dioxide to the lungs, (3) the exchange of the respiratory gases between the cells and the internal environment, (4) the oxidative and other respiratory processes within the tissue cells whereby energy is liberated, and (5) the control of these mechanisms and their correlation with one another and with other bodily processes.

Physiological anatomy

Air entering through the nasal openings is warmed, and some of the grosser impurities are retained by the fine hairs around the nostrils and by the mucous secretion. The nasal cavity just within the external nares (vestibule) is lined with skin. The remaining parts of the nasal cavities are lined with mucous membrane which is covered by a layer of ciliated columnar epithelial cells and scattered "goblet" cells, it is continuous with the membrane lining the accessory nasal sinuses. The nasal lining is very vascular, it contains a venous plexus whose channels anastomose freely and give the mucosa an appearance suggestive of erectile tissue (see nasogenital relationship, chap. 61). The vascular channels are dilated by several conditions, e.g., infections, local irritants, certain anaphylactoid states and a rise in temperature of the inspired air or of the environment, or by local heating of the skin, the mucosa swells and the airway is narrowed in consequence. Cooling the inspired air or skin causes vasoconstriction, as will also the application of adrenaline or ephedrine to the mucosa. The vascularity of the mucosa is also influenced reflexly by the application of heat or cold to the skin of remote regions of the body. The pharynx is, of course, a common pathway for food and air. As the food passes the laryngeal opening this is closed by reflex action and respiration is inhibited. The surface epithelium of the nasal part of the pharynx is provided with cilia and goblet cells. In the oral part of the pharynx the epithelium is of the stratified squamous type. The epithelium of the bronchial tree presents very definite changes as one proceeds from the larynx to the terminal bronchioles. The stratified squamous covering of the upper part of the larynx changes to ciliated in the lower part of the vestibule of this organ. The vocal cords are covered with squamous epithelium, but ciliated cells again line the trachea. The epithelium of the trachea contains also goblet cells and mucous and serous glands. The large bronchioles are similar to the trachea in this respect, but in the bronchioles the goblet cells and deep glands are lost.

CILIA. The nasal secretions are moved toward the nostrils largely by the action of cilia. It is perhaps not generally appreciated that the sinuses are kept clear, under normal conditions, by the beating action of the cilia with which the epithelium is very plentifully supplied. As stated above, ciliated cells are found in the nasal part of the pharynx, in the lower part of the vestibule of the larynx and in the trachea. The cilia beat with a motion which propels material toward the mouth. They become even more abundant when the large bronchioles are reached, but are largely replaced by cuboidal or flattened cells in the respiratory bronchioles.

The efficiency of the ciliated cells of the trachea and large bronchioles in propelling mucus and waste material orally is, under normal conditions, of a high order. The cilia are not influenced by nerve impulses, but are very susceptible to chemical changes in the blood and to substances applied locally. Certain general anesthetics depress their activity and many sedatives exert the same effect. Ciliary action is depressed by cold and increased when the temperature of the cells is raised slightly above normal. The efficiency of the cilia depends in part on the viscosity and stickiness of the material which is in contact with them. Their effectiveness may be varied by changing the properties of this material as well as by an increase or decrease in the rate or force of their beating. The motion of the cilia is wave-like and has been well compared to the undulation of a field of wind swept grain. The individual cilium moves, in the direction in which its force is exerted, with a whip-like motion and then relatively slowly returns to its former position.

ELASTIC TISSUE. The bronchial tree is rich in elastic tissue, most of the fibers being disposed longitudinally in the tunica propria. This elastic membrane, which extends throughout the trachea, bronchi and bronchioles right to the alveoli, is responsible for the recoil of the bronchial tree during expiration and probably in large part for the recoil mechanism of the whole lung. The importance of this elastic recoil will be stressed again later. In the larynx, cartilage supports the special structures necessary for the attachment of the vocal cords and the functioning of the glottis. The cartilaginous rings in the trachea are incomplete on their posterior aspect. This arrangement provides for some contraction of the trachea, but the lumen cannot be obliterated as is the case with the small bronchioles. The tracheal lumen is narrowed somewhat when it is elongated during inspiration.

MUSCLE. The ends of the cartilaginous rings of the trachea may be approximated by the action of the transverse smooth muscle fibers. In the bronchi the bands of fibers tend to become circular and this is seen even more definitely in the bronchioles. The amount of muscle is reduced in the respiratory bronchiole and does not extend beyond this subdivision of the bronchial tree.

BLOOD SUPPLY. The bronchial tree, as far as and including the respiratory bronchioles, is supplied by a rich plexus of vessels derived from the bronchial arteries—branches of the thoracic aorta. The blood is collected by the bronchial veins which in the case of the right lung empty into the azygos vein. Those from the left lung are tributary to the left superior intercostal vein (or sometimes to the accessory hemiazygos vein), but a part of the blood brought to the lungs by the bronchial arteries is returned directly to the pulmonary veins, and some as mentioned below enters the pulmonary vascular bed. The respiratory part of the lung receives its blood from the pulmonary artery, the blood being returned via the pulmonary veins to the left side of the heart. Anastomoses between the pulmonary and systemic systems of vessels occur, however, in the walls of the respiratory bronchioles, so that the blood from this region of the bronchial tree is returned in part to the right side of the heart and in part to the left. Such anastomoses are more numerous in certain inflammatory pulmonary conditions, and it has been thought possible that an abnormal increase in the amount of blood entering the pulmonary vascular bed through such channels might raise the pressure in the capillaries sufficiently to overbalance the colloid osmotic pressure of the plasma and lead to pulmonary edema. It has been suggested that certain types of paroxysmal pulmonary edema may be produced in this way, as a result of reflex vasodilation of the anastomosing channels of the bronchial arteries. However, the greatest quantity of bronchial artery blood which enters the pulmonary system has been calculated to be not more than 1 per cent of the total pulmonary circulation, even with maximal vasodilation. It appears unlikely that this would be sufficient to seriously alter the hydrostatic-osmotic balance in the pulmonary capillaries. Not quite all the blood of the pulmonary artery traverses the capillary bed of human lungs, a portion is shunted through arterio-venous anastomoses, a fact which accounts for the passage of relatively large parasites through the lungs.

NERVE SUPPLY OF THE BRONCHIOLES. Efferent (broncho-constrictor) fibers to the bronchiolar muscle are derived from the vagus, and inhibitor (broncho-dilator) fibers from the sympathetic. Afferent fibers from the lungs run in the vagus. The nerve supply to the pulmonary vessels is considered on page 333. The bronchioles are constricted by pilocarpine, histamine and by certain foreign proteins (anaphylactic reactions). They are dilated by adrenaline, ephedrine and atropine (figs. 29.1 and 34.5, p. 429).

THE BRONCHIOLES AND AIR SACS. Macklin divides the bronchial tree into two parts. The first part which extends from the trachea to the *terminal bronchiole* inclusive, serves simply as an air-conduit and, like the branches and twigs of a tree, possesses no respiratory function. The terminal bronchiole is simply the last of a series of subdivisions of these non respiratory

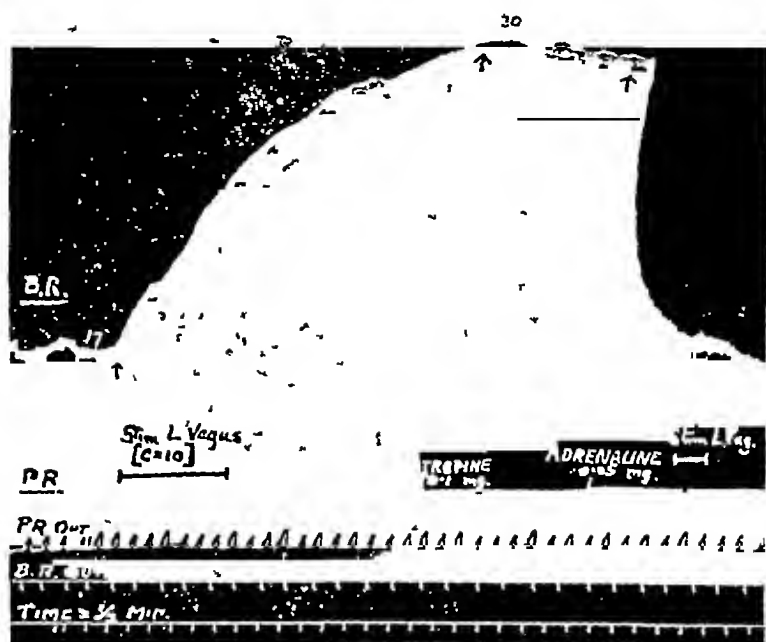


FIG. 29.1 Guinea-pig Showing the broncho-constrictor effect of stimulating the vagus of the same side. The effect lasts after stimulation has ceased. Atropine in moderate dose had a very slight dilator action, adrenaline in small quantity caused marked bronchodilatation. During this action further stimulation of the vagus had an almost negligible effect (After Thornton)

bronchioles. The muscle in its wall is more highly developed than in any other part of the bronchial tree and when fully contracted exerts a sphincter-like action which can completely shut off the air supply to the air chambers beyond. The structures lying distal to the terminal bronchiole are the "leaves" of the bronchial tree. They are respiratory in function, i.e., an interchange of gases between lung air and blood occurs across their walls. This part consists of (a) the *respiratory bronchioles*, (b) *alveolar ducts*, (c) *alveolar sacs*, and (d) *pulmonary alveoli*. The cluster formed of these structures together constitutes a *lung-unit* or *primary lobule* (i.e., the group of structures which like the nephron or renal unit carries out the specific function of the organ). It is the distensible or bellows part of the lung.

The *respiratory bronchiole* has the same diameter as the terminal bronchiole, of which it appears as a branch or a continuation. The *alveolar ducts*, five or six in number, arise from each respiratory bronchiole (fig. 29.2). Each alveolar duct after a variable number of rebranchings gives rise to from 3 to 6 dilatations, the *alveolar sacs*. The bays in the walls of the latter constitute the *pulmonary alveoli*, which are lined by a single layer of flattened epithelial cells cemented together. The alveolar walls contain elastic fibers and a rich network of capillaries. Frequently a single capillary channel alone intervenes between the walls of adjacent alveoli. The blood in the capillaries is therefore separated from the air in the alveoli by two membranes of the utmost delicacy—the alveolar and capil-

lary walls—so the greatest freedom is afforded for the diffusion of gases from the blood to the alveolar air and from the alveolar air to the blood (Willson, Macklin).¹

The bronchioles, as they are traced toward the periphery of the lung, branch and rebranch repeatedly, diminishing in length with each subdivision. The first branchings are about 1.5 mm in length and from 0.3 to 0.4 mm in diameter. The terminal and respiratory bronchioles are from 0.2 to 0.5 mm in length, but of about the same diameter as the earlier subdivisions. That is, the bronchioles, though becoming shorter, show practically no decrease in diameter as they pass toward the periphery. The alveolar sacs, however, are considerably wider than the respiratory bronchiole or than the alveolar duct from which they arise. The diameter of each pulmonary alveolus which has a semiglobular form is from 0.075 to 0.125 mm and the total number in the lungs has been estimated by Zuntz at 750 millions. Willson estimates the total epithelial surface of the lungs at 70 square meters. Of this, probably 55 square meters is respiratory, this is over 25 times the surface area of the skin (p. 618).

¹ The results of Joselyn's studies on the anatomy of the alveoli strongly suggest that the alveolar lining may be a discontinuous membrane and that the capillaries are uncovered. In certain areas, therefore, the air in the alveoli would be separated from the blood only by the capillary endothelium and a small amount of fluid.

THE EXPANSION OF THE LUNGS AT BIRTH

Before birth the alveoli contain a small quantity of fluid, the thorax is unexpanded and completely filled by the quite airless lungs. At this time the smaller proportion of the blood of the right heart passes through the lungs. The remainder passes by the ductus arteriosus into the aorta and via the foramen ovale in the interauricular septum to the left heart. Respiratory movements are made by the fetus *in utero*, a fact clearly demonstrated by the work of Barcroft and of Snyder and Rosenfeld. In man, they can be induced by suitable stimulation after the 17th week. India ink injected into the amniotic sacs of rabbits was found in the alveoli. The significance of the latter observation is debatable. Windle states that he has observed the aspiration of amniotic fluid into the alveoli of experimental animals only under asphyxial conditions. Davis and Potter, on the contrary, injected thorotrast (a radio-opaque material) into the amniotic sacs of ten women about to be delivered by Caesarian section and were able to demonstrate the presence of the material in the lungs of half of the new born infants.

The fetal respiratory movements are readily induced by tactile stimulation. Either anoxia or carbon dioxide excess alone causes strong movements. Asphyxia, caused by clamping the umbilical cord, brings about gasping respirations. The fetal respirations are depressed by anesthetics, narcotics, or by carbon dioxide deficit. In experimental animals the sensitivity to anoxia and CO_2 (inhaled by mother) has been observed to

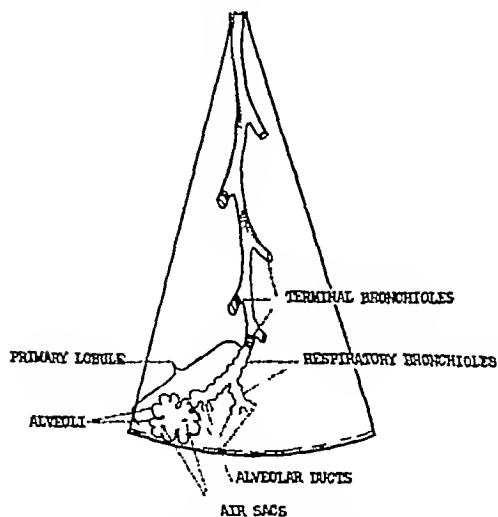


FIG 29.2 Description in text.

vary with the age of the fetus.² Barcroft and Karvonen found that anoxia, induced in sheep fetuses by cyanide, had no effect upon the respirations, except between the 58th and 91st days, before and after this time, no effect was observed. They believe that the effects are brought about through the carotid body (ch 33). The commencement of breathing at birth they attribute to cutaneous nerve stimulation, rather than to mild asphyxia.

At the moment of birth the respiratory movements become more forceful, the diaphragm descends and the external intercostal muscles contract with the result that the diameters of the thoracic cavity are very considerably increased (p 349). A large proportion of the venous blood is now conveyed through the lungs. The general enlargement of the capacity of the thorax—a closed cavity—tends to reduce the pressure on the outer (pleural) surfaces of the lungs. The greater the degree of enlargement of the chest, the greater will be the reduction in the pressure upon the outer pleural surfaces. The interior of the lungs, however, is in direct communication, through the air passages, with the atmosphere. The visceral and parietal pleurae being inseparable the lungs follow the thoracic wall as it enlarges, and therefore must expand. The rarefaction of the pulmonary air as a result of the expansion results in a flow of atmospheric air into the lungs. Full expansion of the lung is not attained until some few days after birth. The lung throughout the individual's life remains in the expanded position—pressed as it were against the thoracic framework as a result of the greater pressure exerted upon the alveolar than upon the pleural aspects of the pulmonary tissue (p 347).

Rhythmically alternating increases and reductions of the expanded state of the lungs initiated at birth continue throughout life and constitute respectively the *inspiratory* and *expiratory* phases of respiration. The alternate inflations and partial deflations of the lung are the direct result of corresponding changes in the capacity of the thoracic cavity occasioned by the movements of the diaphragm and other respiratory muscles. Changes in pressure within the lung—the *intrapulmonary pressure*—and upon its pleural surfaces—the *intra-*

² Windle has drawn attention to the serious effects upon the central nervous system of fetal anoxia, even of short duration, and suggests that asphyxia of the new-born, even though of short duration, may be responsible in some instances for nervous and mental conditions, e.g., subnormal mentality, or even feeble-mindedness, in after life.

pleural pressure—occur coincidently with the alterations in lung volume

INTRAPULMONARY PRESSURE In the resting position of the chest the intrapulmonary pressure is atmospheric, but it varies rhythmically with the phases of respiration, rising above atmospheric pressure during expiration and becoming subatmospheric during inspiration. These variations may be demonstrated by connecting one nostril with a manometer and breathing with the mouth closed. The pressure will be found to be about -2 mm Hg during the inspiratory phase and to rise to $+3$ or $+4$ mm Hg during the expiratory phase of ordinary quiet respiration.³ The variations are accentuated considerably during forced respiration. The maximal negative pressure capable of being developed within the lungs by a forced inspiration, as when a strong sucking effort is made, is from -40 to -50 mm Hg. When expiratory efforts are made against a closed glottis, as in coughing, during muscular effort with straining, or during defecation or micturition the intrapulmonary pressure becomes raised by from 10 to 40 mm Hg above the pressure of the atmosphere. If the free flow of air into and out of the lungs is hindered as a result of some diseased condition the intrapulmonary pressures will be increased beyond the normal range.

THE PLEURAL CAVITIES The lungs are invested by the visceral layer of the pleural membrane. The membrane is reflected from the root of each lung on to the inner aspect of the walls of the chest and upper surface of the diaphragm—this is the parietal layer of the pleura. The two layers thus form a closed membranous sac on each side of the chest. The potential space enclosed by the pleural membranes is spoken of as the pleural cavity. In health *no actual space exists, the two membranes are in apposition* except for a thin film of fluid which serves as a lubricant to allow the surfaces to glide over one another during the respiratory movements. This potential cavity may, however, as the result of disease become an actual one. Serous fluid (hydrothorax), pus (pyothorax or empyema), blood (hemothorax) or air (pneumothorax) may collect and separate the two layers. Between the two pleural compartments lies the *mediastinum*, a space which is subdivided by the heart with its pericardial investment into an anterior and a posterior part—the *anterior* and *posterior mediastina*.

A simple puncture of the pleural cavity (thorac-

centesis) is sometimes followed immediately by fainting or collapse of the subject and may prove fatal. This so-called pleural shock has been attributed to air embolism but is most likely due to a pleural reflex which brings about cardiac slowing and a fall in blood pressure, being in this respect similar in character to the carotid sinus reflex.

INTRAPLEURAL PRESSURES It has already been mentioned that the pressure on the pleural surfaces of the lungs is less than that upon their alveolar surfaces, i.e., the intrapleural pressure is subatmospheric. We must now consider the manner in which this "negative" pressure is produced. As stated above (p. 346) when the chest cavity is first expanded, the lungs are carried outwards by the inflow of air to fill the enlarged space. If this were all that occurred, the pressure within the lung and in the pleural cavity would be equalized and in the expanded position of the thorax after birth, as in the unexpanded state in the unborn animal, the pressure in the pleural cavity would not be subatmospheric. The expansion of the thorax, however, and the consequent inflation of the lungs puts the pulmonary tissue upon the stretch. In other words, the closed thoracic box, as a result of the first breath, becomes too large for the lungs to fill by a simple unfolding and distention of the walls of the air spaces. The elastic tissue of the bronchial tree, blood vessels and of the air sacs themselves is put under stress and is constantly pulling against the stretching force. This pull or recoil of the elastic lung amounts, in the adult when the chest is about midway between inspiration and expiration, to a pressure of from -4 to -5 mm of mercury. The existence of such a pressure can be demonstrated indirectly by connecting a manometer with the trachea of a dead subject and puncturing the chest wall. Thus the negative pressure in the pleural cavity is abolished and the lungs are permitted to recoil, i.e., to collapse. Air is expelled from the alveoli and the manometer registers a pressure of $+4$ mm Hg. This represents the pull which had been exerted before the puncture was made and is just equal in amount to the negative (suction) pressure in the pleural cavity. In the new-born the lungs fill the thoracic cavity with comparatively little stretching. The distension of the lungs increases, however, in later years since the thoracic cage grows more rapidly than the lungs, the elastic pull in consequence also increases and with it the intrapleural negative pressure. If, in the foregoing experiment, the

³ As a result of the obstruction to breathing offered by the apparatus, these values are somewhat greater than actually exist during normal breathing.

lungs are distended maximally before the thoracic puncture is made, i.e., if they are fully stretched, then the pulmonary recoil causes a rise of 30 mm Hg or so in the manometer column. The intrapleural pressure may be demonstrated directly by plunging a cannula connected with a manometer into the pleural cavity in such a way as to prevent leakage between it and the margins of the puncture. The manometer registers a negative pressure equivalent in amount to the positive pressure recorded in the previous experiment. In other words, the mercury is "sucked" toward the pleural cavity until the pressure within the latter just equalizes that of the atmosphere.

Another and apparently a more important factor in preventing the separation of the pleural surfaces, and maintaining the lungs normally in the expanded state against their own elastic pull, is the "hydraulic traction" (West) exerted by the film of fluid between the layers of the pleura. This so-called hydraulic traction depends upon the tensile strength of water (as is exhibited when an attempt is made to pull apart two wet, smooth surfaces, e.g., two apposed moist glass slides). A film of water is capable of withstanding a pull of 3600 mm. Hg per square centimeter. The introduction of air into the pleural cavity permits shear to take place, just as the two glass slides can be easily separated if a small quantity of air is allowed to enter between them. That inequality of pressure on the two sides of the two aspects of the lung is not the sole, nor perhaps the most important, factor holding the pleural layers in close apposition is evident from the fact, pointed out by Burns, that the intrapulmonary pressure can be reduced far below the intrapleural pressure without the lungs collapsing.

The intrapleural pressure (and the pressure throughout the thoracic cavity generally) is always subatmospheric under ordinary circumstances,⁴ even after death. This "negative" pressure is increased during inspiration—since then the distention of the elastic lungs is greater—and reduced during expiration. During the former phase of quiet respiration (human) it amounts to about -6 mm Hg, during an ordinary expiration is about -2.5 mm Hg. In the midposition as stated above it is about -4.5 mm Hg. When the movements are

forced it may be very greatly increased or diminished in the respective respiratory phase. During a strong inspiratory effort with the closed glottis it may amount to -40 mm Hg and in forced expiration under the same circumstances it is abolished and a positive pressure of 50 mm Hg or so substituted. These changes in intrathoracic pressure exert an influence upon other thoracic structures. An increase in the "negative" pressure causes the thin-walled veins and auricles to expand and fill with blood drawn from extrathoracic regions (fig. 29.3). On the other hand, in forced expiration against the closed air passages the thoracic walls press powerfully upon the air-filled lungs. The rise in pressure which results is transmitted to structures lying in the mediastinum. Blood is thus expelled from the large intrathoracic veins and auricles into the veins of the abdomen and neck.

THE RESPIRATORY MOVEMENTS

What has been said in the foregoing paragraphs should have made it clear that the flow of air into and from the lungs depends entirely upon changes in the capacity of the thoracic cavity. The lungs play a purely passive rôle. The air is not drawn in and expelled by active dilatation and contraction of the pulmonary passages, as was the belief at one time. Air is drawn in or forced out strictly in accordance with the pressure differences between the atmosphere and the lung air caused by the expansion or contraction of the thoracic boundaries, i.e., as air is drawn into and expelled from a bellows. The principles are well illustrated by the model shown in figure 27.10. We will now consider how these changes are brought about.

The respiratory movements of an adult person occur normally at the rate of from 16 to 18 double excursions (inspiration and expiration) per minute. In the new-born infant during quiet breathing the rate is between 30 and 40 per minute.

During inspiration the thoracic cavity is enlarged in all diameters, vertical, anteroposterior and transverse. The enlargement, however, is not equal in all directions. The upper part of the thorax increases much less in capacity than does the lower part, and since the position of the spinal column remains relatively fixed the increase in the anteroposterior diameter of the thorax is due mainly to an expansion forwards. The increase in the vertical diameter is due, not to an upward expansion of the chest cavity but to the downward elongation resulting from the descent of the diaphragm.

⁴ The intrapleural pressure may be measured in the human subject by inserting a hollow needle into the pleural cavity, injecting a small quantity (40 cc. or so) of air and connecting the needle with a water manometer and a recording system.

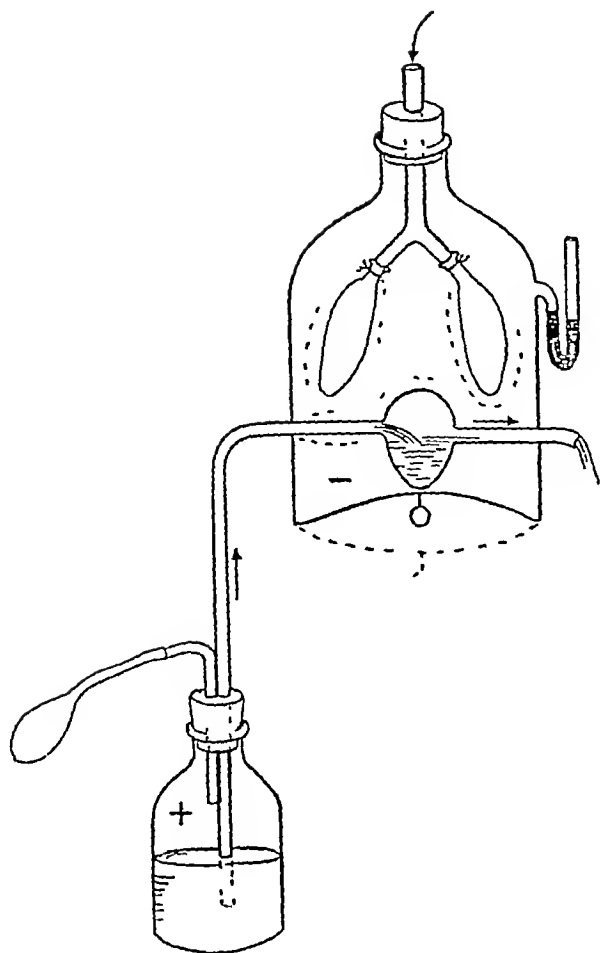


FIG 29 3 Model to illustrate the manner by which changes in thoracic capacity cause corresponding changes in the volume of air in the lungs and affect the return of blood to the heart. The large glass chamber represents the thorax, it is hermetically sealed and has a flexible bottom or diaphragm. The Y-tube, which is in communication with the atmosphere, represents the trachea and bronchi, the lungs are represented by the attached balloons composed of thin rubber. The pressure within the chamber, i.e., surrounding the balloons, is subatmospheric to start with, the balloons are therefore partially expanded. When the diaphragm is drawn down (as indicated by the dotted lines) the pressure within the chamber is further reduced. The rubber balloons are distended to a corresponding extent by atmospheric air entering through the Y-tube. As the diaphragm is allowed to rise again, the "negative" pressure within the chamber returns to its previous value and the elastic balloons recoil to their original dimensions.

The bottle shown in the lower left-hand part of the drawing contains fluid upon which pressure can be exerted to cause a steady flow up the tubing into the small oval chamber within the larger one. The oval chamber may be taken to represent the heart, the tubing connecting it with the pressure bottle represents the large veins, and the tubing leading from its right side the arteries. If the tubing on the left, as well as the upper part of the oval chamber (which corresponds to the auricles) be composed of some thin resilient material, then during the inspiratory reduction in intrathoracic pressure they will undergo expansion (dotted lines). A greater body of fluid will

Unequal enlargement of the thoracic box entails unequal expansion of the lungs. The lung is not distended equally from a center as in the inflation of an elastic-walled globe. Keith distinguished three zones in the expanding lung.

(a) A non-expansile *root zone* containing the bronchus, pulmonary vessels and lymphatics and their main divisions.

(b) An *intermediate zone* in which the vascular and bronchial branches radiate outwards toward the lung surface. Between these rays lies expansile pulmonary tissue. This zone therefore consists of tissue of varying degrees of expansibility, that lying near the periphery of the rays being more expansile than that situated more centrally.

(c) An *outer or subpleural zone*, from 1 to 1½ inches deep, of maximal distensibility.

Those regions of the lung lying in relation to the relatively immobile regions of the thoracic walls, namely (a) the dorsal surface of the lung apex, (b) the posterior surfaces of the lungs in contact with the spinal column and attached segments of the ribs, and (c) the mediastinal surface lying in relation to the pericardium and other structures of the mediastinum, are expanded *indirectly*. The parts of the lung which are *directly* expanded during inspiration are those lying in contact with the freely movable boundaries of the thorax, namely (a) the sternum and ribs and (b) the diaphragm.

It is evident that those portions of lung in contact with practically stationary regions of the thoracic walls can only be expanded indirectly, that is, when other parts of the lung move out of the way. This could not occur did the root of the lung remain fixed. As a matter of fact the lung root moves downwards, forwards and laterally during inspiration (fig 29 4) and, as shown by Macklin by X-ray studies upon human subjects, the bronchial tree becomes elongated (stretched) during the inspiratory phase. The trachea becomes stretched and the apex of the lung actually descends as it expands. During expiration the highly elastic bronchial tree recoils to its previous length and the lung root ascends (fig 29 5). If the root of the lung were fixed, little expansion of a region such as the apex or of other regions classed

in consequence be transferred from the bottle to the small chamber representing the heart (see p 168).

The manometer inserted into the wall of the large chamber registers the pressure changes (indicated by dotted lines) occurring during the descent and ascent of the flexible diaphragm.

as expanding indirectly could result. Nor could anything but a very moderate expansion of other regions (e.g., costosternal and diaphragmatic, etc.) occur if the bronchial tree were unable to lengthen. It is the *elongation* of the rays as described above rather than the widening of the spaces between them at their original lengths, like the separation of the sticks of a fan, that is of importance in permitting the expansion of the intervening pulmonary tissue.

The effect of the enlargement of the thorax is exerted first and to the greatest extent upon the lung tissue in relation to the movable parts of the chest walls. The inspiratory decrease of intrapleural pressure in the diaphragmatic regions of the thorax is considerably greater than that in the region of the apex or in other parts of the lung which are expanded indirectly. As a result of the greater negative pressure in the lower part of the thorax a horizontal groove is sometimes developed here (Harrison's sulcus) when the framework, as in rickets, is soft and yielding. The restricted expansion of the air sacs of the apex and other regions of the lung which are expanded indirectly has been held responsible for their

being so commonly the primary site of tuberculous infection.

THE ENLARGEMENT OF THE THORACIC CAVITY DURING INSPIRATION

This is effected, according to Keith, by four distinct mechanisms which consist of the movements of

(a) The *thoracic lid* or *operculum* (1st rib and manubrium sterni)

(b) The *upper costal series* (2nd to 5th ribs inclusive)

(c) The *lower costal series* (6th to 10th ribs inclusive) and the *diaphragm*

(d) The *floating rib series* and the *muscles* of the abdominal wall

(a) The thoracic lid or operculum

The thoracic lid or operculum is formed by the first pair of ribs and the manubrium sterni. It is

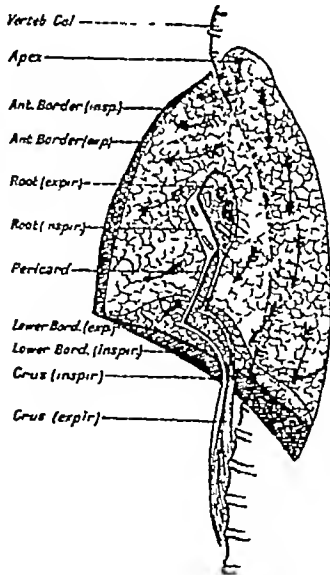


FIG 294 Mediastinal aspect of the right lung to show the respiratory movement of the lung root. The crus of the diaphragm is also indicated, and its attachment to the root of the lung through the pericardium. The arrows indicate the direction of the inspiratory movement of the various parts of the lung. (After Keith.)

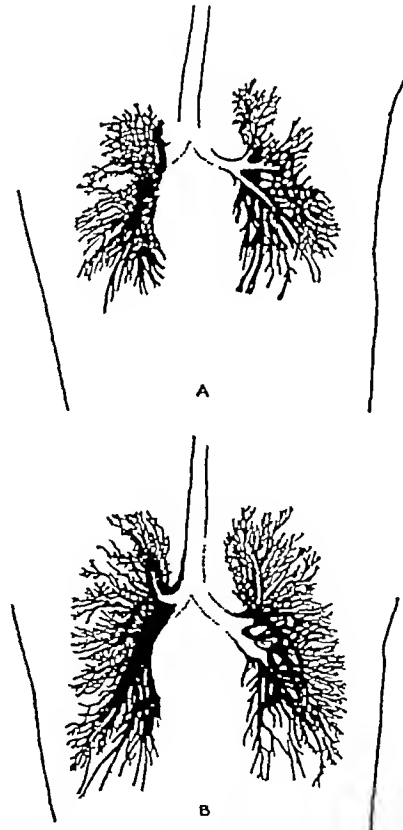


FIG 295 X ray photographs (retouched) of the bronchial tree of a young woman. A, in full expiration, B, in full inspiration. (After Macklin.)

jointed behind to the spinal column and in front to the sternum by the manubrio-sternal joint. During the elevation of the thorax in inspiration the thoracic lid moves as a single piece upon the body of the sternum, assuming a more horizontal position (by from 1° to 16°). That is, the manubrium is pushed upward and forward (fig 29 6). Thus, the upper part of the thorax is increased in its anteroposterior diameter. The anterior portion of the lung apex is directly expanded to some extent by this mechanism. The extent of the movement of the thoracic lid varies considerably in different individuals and with the depth of inspiration. It is very slight in quiet breathing. The manubrio-sternal joint becomes ankylosed in later life but rarely before the 60th year.

(b) The upper costal series

The 2nd, 3rd, 4th, 5th and 6th ribs slope obliquely from behind downwards and forwards. Each rib is longer, its direction more oblique and it makes a fuller sweep outwards than its neighbor immediately above. During inspiration these ribs (with the exception of the 2nd) assume a more horizontal position, their anterior portions moving upward and forward. That is, each rib rotates around an oblique horizontal axis parallel to its neck (fig 29 8CD). The sternum is thrust forward and upward, executing a movement at the manubrio-sternal joint. These movements increase the anteroposterior diameter of the thorax. The elevation of the ribs is effected by the external intercostal muscles. The muscle fibers pass obliquely downwards and forwards from the lower border of one rib to the upper border of the rib below. When the muscle contracts it exerts a pull upon these attachments which tends to depress the upper rib of the pair and to raise the lower. The first rib, however, acts through the contraction of the scalene muscles, as a fixed point above, so that contraction of the external intercostals can only result in an elevation of the ribs. Owing also to the obliquity of the fibers which are attached below to the anterior end of the long arm of a lever and above to the posterior end of the long arm, a distinct mechanical advantage is given to the upward movement (fig 29 7). The internal intercostal muscles are, in the cat at any rate, as shown by Bronk and Ferguson, expiratory in function. The internal and external intercostals receive impulses alternately along the intercostal nerves.

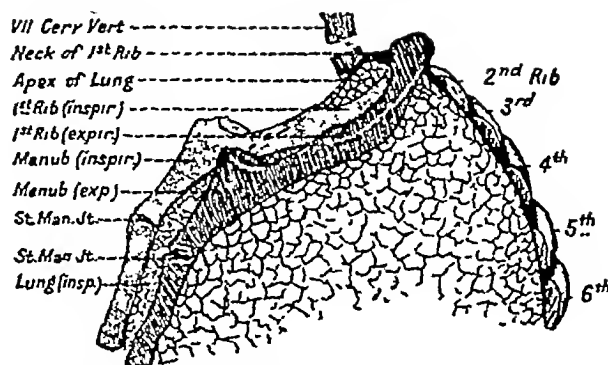


FIG 29 6 Diagram to show the respiratory movements of the first pair of ribs and manubrium sterni, and the effect of these movements of the expansion of the apex of the lung (After Keith)

The strongly bowed mid-portion of the body of each rib from the 2nd to the 6th also becomes elevated, in relation to its two ends, rotating around an oblique anteroposterior axis. This movement, which is compared to the raising of a bucket-handle to a more horizontal position, increases the transverse thoracic diameter (fig 29 8AB).

(c) The lower costal series and the diaphragm

The ribs from the 7th to the 10th also swing outwards and upwards (bucket-handle movement) during inspiration, rotating around an oblique anteroposterior axis which passes through the mid-line in front and the necks of the ribs behind. The subcostal angle is widened by this movement and the transverse diameter of the lower part of

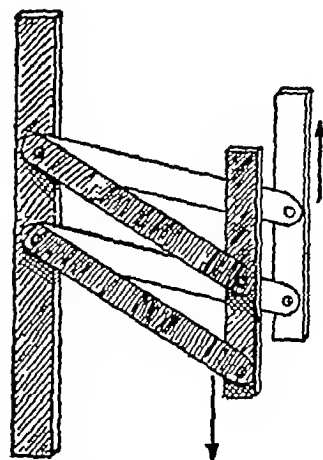


FIG 29 7 Diagram to illustrate the action of the external intercostal muscles (EI) during inspiration, EXP, expiration, INSP, inspiration

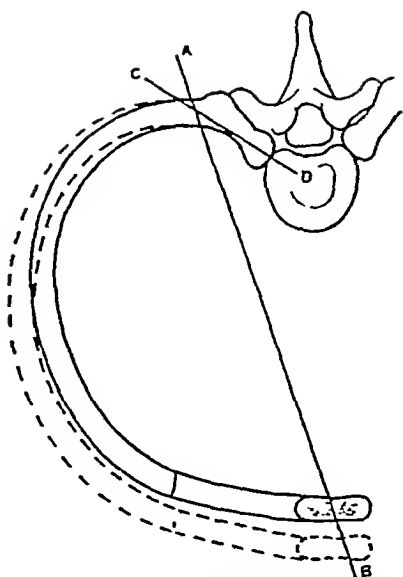


FIG. 29.8 A diagram showing the axis of movement (AB and CD) of the ribs from the 2nd to the 6th. The ribs from the 7th to the 10th also make a movement around an antero-posterior axis but not around the axis CD. The interrupted lines indicate the position of the rib in inspiration. (After Gray.)

the thorax increased, the antero-posterior diameter is slightly reduced.

The diaphragm is the chief muscle of respiration, its movements being responsible during deep breathing for 60 per cent of the total amount of air breathed. It consists of a musculo-tendinous sheet arched toward the thoracic cavity. The tendinous portion is centrally placed (central tendon) and is adherent to the pericardium. The muscular tissue is placed circumferentially. The diaphragm consists of two parts which differ from one another in their actions.

(a) The *costosternal* part arises from the back of the xiphoid and the cartilages and adjacent portions of the six lower pairs of ribs. It is attached to the anterior edge of the central tendon.

(b) The *lumbar* or *crural* part arises from the fibrous arches over the quadratus lumborum and psoas muscles and by two fleshy bundles (the crura of the diaphragm) from the bodies of the upper lumbar vertebrae. These fibers are inserted into the posterior margin of the central tendon.

ACTION OF THE DIAPHRAGM. The diaphragm descends during inspiration and ascends during expiration. In full expiration its upper limit lies at a level situated between the costal cartilages of

the 4th and 5th ribs. In quiet breathing the range of its movement is about 1.2 cm and in forced breathing about 3.0 cm. The total diaphragmatic surface is about 270 sq. cm. A descent of 1.0 cm therefore (assuming that all regions descend practically to the same extent) will increase the thoracic capacity by 270 cu. cm and cause a corresponding volume of air to enter the lungs.

As the diaphragm descends its domed shape alters very little, it may be seen by means of the fluoroscope to move up and down like a piston (fig. 29.9). At the end of expiration a considerable proportion of the diaphragmatic surface is in contact with the chest wall as high as the 6th or 7th rib, but during inspiration it is "peeled off" the thoracic boundary, while the base of the lung expands to fill the space (pleural sinus). As a result of the slight indrawing of the intercostal spaces caused by this movement, a faint shadow may be seen to move down the side of the chest wall in most normal persons. This is known as Litten's sign. The *costosternal* part of the diaphragm, (using the lower ribs, which through the action of the external intercostals, serve as fixed points) moves downward and forward, depressing the abdominal viscera. Thus, the capacity of the lower part of the thorax is increased. The abdominal wall distends but when, as a result of the resistance offered by the abdominal muscles, the downward movement of the viscera becomes arrested, the latter act as a fixed point for the continuing contraction of this part of the diaphragm, its force is now spent in raising the lower ribs to which it is attached. Through this action the sternum is thrust forward and upward. The *spinal* or *crural* part in its descent acts solely in increasing the vertical diameter of the thorax.

The excursions of the diaphragm and consequently its mid-position as well are influenced by (a) the upward pull of the subatmospheric intrathoracic pressure, and (b) the abdominal viscera. In the standing position the weight of the latter exerts a downward pull and so aids the descent of the diaphragm but hinders its ascent, the mean or mid position of the diaphragm is therefore taken up at a lower level than in recumbency when the viscera exert an upward pressure. (c) The abdominal muscles, these, when lax and the body in the standing position, allow the viscera to subside to a lower level and so to increase the downward pull upon the diaphragm. In persons with extremely weak abdominal muscles, such as sub-

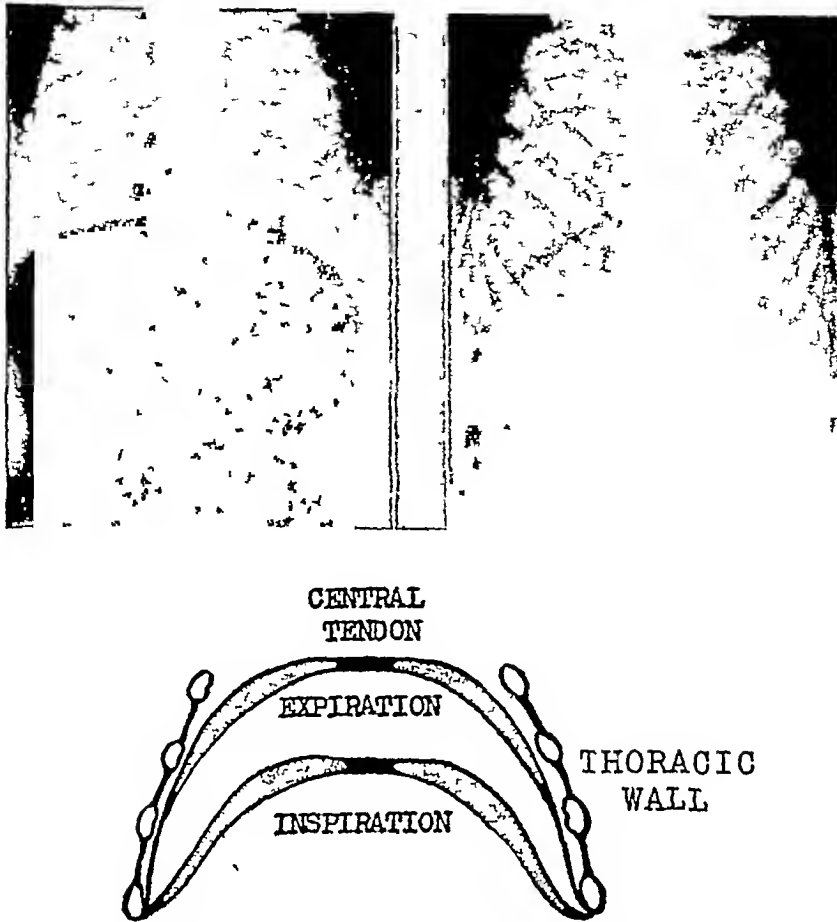


FIG 29.9 Upper, radiogram showing the position of the diaphragm during expiration and inspiration. Note the effect upon the position and shape of the heart. (After Norris and Landis.) Lower, diagram showing the expiratory and inspiratory positions of the diaphragm.

jects of visceroptosis, the downward pull upon the diaphragm greatly interferes with its movements. The breathing is then largely costal.

(d) *The floating ribs (11th and 12th) and the abdominal muscles*

Functionally the floating ribs must be considered with the abdominal muscles which are the antagonists of the diaphragm. The recti and oblique muscles relax as the diaphragm descends, they contract with its ascent.

EXPIRATION

Expiration is to a large extent a passive movement, in quiet breathing it is probably entirely so. That is, the contraction of the inspiratory muscles ceases, the thoracic framework tends through its own weight and inherent elasticity to resume its former position, the elastic lungs recoil and the relaxed diaphragm is drawn upwards toward the thoracic cavity by the "negative" intrathoracic pressure which is greatest at the end of inspiration.

There is, however, also a definitely active element in a forced expiratory movement. As mentioned above, the abdominal muscles contract and by pressing upon the viscera aid the ascent of the diaphragm, i.e., the diaphragm is "pushed" up by the increased intra-abdominal pressure resulting from the contraction of the abdominal muscles, as well as "sucked" up by the subatmospheric pressure within the thorax.

During forced expiration the internal intercostals whose fibers, like those of the external intercostals, course obliquely but in the opposite direction (downwards and backwards) contract and so aid in the depression of the ribs. The restoration of the thorax to its previous diameters is, of course, accompanied by a corresponding reduction in the capacity of the lungs and the expulsion of air from its air spaces.

MOVEMENTS OF THE BRONCHIAL TREE DURING RESPIRATION In addition to the inspiratory elongation of the bronchial tree mentioned above, the bronchioles and smaller bronchi dilate during

inspiration and constrict during expiration. These rhythmical bronchial movements appear, from the studies of Ellis and of Nicholson, to be purely passive in nature and not integrated through reflex mechanisms with inflation and deflation of the lungs. The bronchioles also exhibit a peristaltic movement which can be detected by X-ray photographs of the bronchioles after the injection of lipiodol* or other material which is opaque to the X-ray. It is not thought that the peristaltic movement plays any part in the movement of air, but it appears to assist in the movement of foreign material towards the larger air tubes. This bronchiolar peristalsis is said to be increased in lung abscess and diminished in bronchiectasis. It is decreased by morphine.

PULMONIC ALVEOLAR VENTS. The existence of certain pores in the alveoli (pores of Kohn) have long been recognized, but the difficulty of being certain that they are not artefacts produced by the method of preparing the sections has prevented their wide acceptance as normal structures. Macklin, however, has been able to demonstrate the presence of such vents in thick sections prepared from the lungs of many species including man. Examination of the sections justifies the belief that they are normal communications between the alveoli. While the function of the vents is not as yet completely understood, it appears probable that they play a significant part in the equalization of pressures in groups of the alveoli, particularly perhaps during forced inspiration. The vents are opened wide during inspiration and may be entirely closed in expiration.

COLLATERAL RESPIRATION. This phenomenon (Van Allen, Lindskog and Richter) depends on the fact that the lobular divisions of the bronchial tree of any one pulmonary lobe are interconnected. Air may pass from one alveolus to an adjoining one presumably through the alveolar vents but possibly also by rapid diffusion. This communication may be demonstrated when a bronchus supplying one lobule is obstructed, for it is then found that the alveoli of the obstructed lobule may be well ventilated for prolonged periods. Collateral respiration is prevented by inflammatory exudates or secretions or in circumstances in which the alveolar walls are not adequately distended. In this latter case the vents may not be open. In species in which the interlobular septa are complete as in man there is no provision for collateral respiration between different lobules. Experimental evidence has been obtained which indicates that atelectasis of a lobule does not develop after obstruction of its bronchus. Conversely, re-aeration of an atelectatic lobule may

take place by this collateral route. It would thus appear that considerable anatomical and physiological evidence has been obtained for a peripheral mechanism which has as its objective the adequate and uniform but not excessive distension of the pulmonary alveoli.

EXPULSION OF FOREIGN MATERIAL—THE ACT OF COUGHING PHAGOCYTOSIS. There are three mechanisms for the expulsion of foreign material from the larynx, trachea and lower air passages as far as the terminal bronchioles. First, the action of the cilia, second, the peristaltic motion of the bronchioles, and third, the cough reflex. The peristaltic movement has been referred to above. The cough reflex is most commonly initiated by the stimulation of afferent nerve endings in the region of the tracheal bifurcation, the most sensitive area, or in the laryngeal mucosa. It may also be initiated from the excitation of vagal afferents in the lungs, or from nerve endings in the pleura. Ear disease, through the stimulation of terminals of the auricular branch of the vagus (Arnold's nerve), may also cause coughing. The act itself consists of a short inspiration followed immediately by closure of the glottis and a forcible expiratory effort. A considerable degree of pressure is thereby developed within the lung. The glottis then opens suddenly and offending material is moved a variable distance along the air passage. If it reaches an insensitive area the coughing ceases. During the subsequent inspiration the irritating particle, if not large enough to seriously obstruct the air passage, remains during the subsequent inspiration in its new position, from which it is carried forward again during succeeding expulsive efforts until it is swept away from sensitive areas.

Inert, non-irritating foreign matter of a particulate nature, e.g., dust or carbon particles, which has entered the alveoli, is removed by large ameboid cells—the alveolar macrophages, or "dust cells." Normally, a few of these cells are to be seen in the alveoli, but they are attracted in large number by the presence of foreign particles. The origin of these cells is a controversial question. Some are thought to originate from the monocytes of the reticulo-endothelial system, others from cells in the alveolar septa (septal cells). The foreign material is conveyed by the macrophages into the lymph channels draining the alveoli. In city dwellers or those who work in dust laden atmospheres, e.g., coal miners, the lymph nodes are dark with the scavenged material. Irritating matter, espe-

* A preparation of iodized oil.

cially pathogenic bacteria, are attacked as in other situations by the neutrophilic phagocytes of the blood

Hiccup or singultus is a spasmodic and purposeless contraction of the diaphragm which results from many causes or may occur without known cause, it occurs rarely in epidemic form. It is usually reflex in nature being initiated by some abnormal stimulation of the afferent nerve terminals in the diaphragm. The fibers of the phrenic nerve constitute the efferent limb of the reflex. In some instances, hiccup may be due to stimulation of the respiratory center itself by some agent in the blood. Hiccup following abdominal operations may be most intractable and endanger the life of the patient. The inhalation of an air mixture containing 6 or 7 per cent of carbon dioxide (or simply breathing and rebreathing from a bag for a number of respirations) has been reported to be a valuable means of terminating an attack.

Artificial respiration

Schafer's prone pressure method is the most widely practised means of artificially ventilating the lungs when spontaneous respiration has failed, and, all things considered, is probably the most valuable. It has the advantage that it requires no special equipment and can be undertaken by one person, artificial respiration can, therefore, be started immediately, which is of the utmost importance.⁶ The operator kneels astride the subject stretched prone beneath him and with the palms of his hands upon the subject's lower ribs leans forward and exerts gentle but firm pressure for 2 seconds. He then straightens up and releases the pressure for 2 seconds. These alternating movements are repeated at the rate of normal respiration. The downward and forward movement compresses the thorax and pushes the diaphragm upward, thus expelling air from the lungs. The recoil of the thorax to the resting position follows the release of the pressure, and thus induces artificial inspiration. In a normal apneic subject the volume of tidal air, when this method is used, is about the same as or a little below that of the subject breathing naturally, the pulmonary ventilation appears to depend upon the tonus of the respiratory muscles, so long as the nerve

centers which govern it are viable, rather than upon the degree of force exerted by the operator (Henderson). In other words, the chest after compression expands again to a certain volume determined by the tone of the diaphragm and intercostal muscles. It has been found in experiments upon animals that artificial respiration, by compression and decompression of the chest, immediately and for some minutes after death induces a degree of pulmonary ventilation equal to that during normal breathing. But after the lapse of 10 minutes or so the volume of tidal air is reduced to zero. Meltzer has also shown that the pulmonary ventilation effected by this method of resuscitation is much reduced in animals in which the respiratory muscles have been paralyzed by curare. Greater ventilation of the lungs than in Schafer's method has been claimed for the *Neilson-Schafer* method as modified by Drinker, and for *Eve's rocking method*.⁷ In the Neilson-Schafer-Drinker procedure the extended arms of the subject are raised above his head by a second operator during the inspiratory phase of "Schafer respiration." This helps to expand the chest. The arms are returned to the ground during the expiratory phase. In the rocking method the inspiratory and expiratory movements of the diaphragm are brought about by the weight of the abdominal viscera, the subject is placed prone upon a stretcher which is then rocked on a trestle around a transverse axis. The tilt of the body both in the head-down and feet-down positions is from 45 to 50 degrees. Ten double movements are made per minute. This method has been shown to increase the cardiac output, and thus to secure a better blood supply to the coronary and cerebral vessels, and to increase the oxygen uptake from the lungs. The volume of tidal air, during the rocking of a normal person during a period of apnea (induced by forced breathing), is claimed to be greater than that of the same subject breathing naturally at rest. A machine-motored rocking bed has been designed to carry out prolonged artificial respiration as required in cases of paralysis of the respiratory muscles.

The possibility of supplying oxygen intravenously

⁷ A physiologically admirable means of resuscitation, which has been used for centuries and goes back to Biblical times, is that which is now called the mouth to mouth insufflation method. The operator applies his lips to the mouth of the apneic subject, and endeavours with his own breath (containing a high percentage of carbon dioxide), to inflate the patient's lungs.

⁶ Ross, after a review of the literature of resuscitation, states that no instance of revival was found in which more than 15 minutes had elapsed between the cessation of breathing and the commencement of artificial respiration.

has been investigated on several workers, and the experiments of Negovsky indicate that the transfusion of arterial blood, combined with glucose and adrenaline and vigorous artificial respiration, is the most efficient means of resuscitating animals. He found that the heart ceases to beat about the same time as the last natural respiratory movement, and that it was impossible to revive the animal if more than 6 or 8 minutes had elapsed after this time, for irreparable damage to nervous centers had then occurred.

During the last few years several methods have been developed for the long continued application of artificial respiration by various mechanical devices. The work was begun by Thunberg in Lund who devised an apparatus called the barospirator. The subject was placed inside a metal chamber in which the pressure was raised and lowered rhythmically by means of the stroke of a large piston. The interchange of air within the lungs was caused by a rise and fall of pressure of the air in the external atmosphere. This apparatus was effective and a model was built large enough to accommodate patient, nurse and doctor. All three were ventilated without movement of the chest. A certain amount of discomfort was experienced in this cabinet due to the change in pressure on the two sides of the eardrum during the increase and decrease of air pressure. A more generally applicable model was produced by Drinker at Harvard. In this type the patient's head remains outside the cabinet (fig. 29 10). The chest is expanded by reducing the pressure within the cabinet and as the pressure is raised again, the natural elasticity of the lung causes expiration. Forced expiration, however, may be produced by raising the pressure above atmospheric. Patients have been adequately ven-

tilated with this apparatus for many months. Some difficulty is occasionally encountered in the regulation of the rate and depth of respiration, but by determining the oxygen saturation of the arterial blood or watching for signs of cyanosis an observer can usually regulate the ventilation quite satisfactorily. Artificial respiration for prolonged periods by Eve's method has been mentioned.

Two other types of apparatus should be mentioned, (1) the rubber model of Sahlin which operates like the Drinker machine but is applied only to the chest and (2) the Bragg Paul pulso which consists of a neoprene elastic bandage placed around the chest. The bandage, when inflated by an electrically driven bellows, compresses the chest, which returns to the normal position by virtue of its own elasticity during the interval between the compressions.

The pulmotor and other similarly devised resuscitators which force air into the chest and suck it out again as though the lungs were rubber bags, though so appealing in their mechanical efficiency are physiologically unsound. They are not only less effective than other methods but may be actually harmful. One ill effect is the tendency to reduce the carbon dioxide tension of the blood and thus remove the natural stimulus to breathing and to depress the circulation as well. Those machines which inflate the lungs with an oxygen-carbon dioxide mixture, and permit expiration to take place simply by interrupting the air current and thus permitting collapse of the thorax, are much sounder in principle.

The object of artificial respiration is not only to aerate the lungs but also to stimulate the respiratory center. Due largely to the insistence of Yandell Henderson, carbon dioxide is now added,

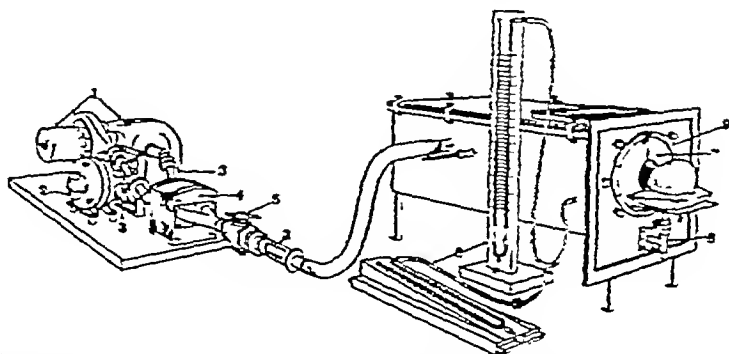


FIG 29 10 The Drinker respirator: 1, pumps; 2, motor; 3, vents; 4, alternate; 5, valves; 6, manometers; 7, external shutters; 8, adjustment for head rest; 9, adjustable ring to hold collar in place. (After Shaw and Drinker Jour. Clin. Investig. 1930/8/33)

whenever possible, to the respired air which should contain from 40 to 50 per cent of oxygen. The carbon dioxide percentage of the air mixture is from 5 to 10 per cent. The high tension of this gas combined with the rhythmical inflation and deflation of the lungs, which presumably cause the discharge of afferent impulses, encourages the return of spontaneous breathing.

Resuscitation of the new-born Asphyxia in the new-born infant is usually due to the failure of the lungs to expand fully, alveoli in areas of the pulmonary tissue remaining in the fetal state (atelectasis). The methods of resuscitation which may be effective in the adult are likely to fail. Cutaneous stimulation, e.g., slapping, hot or cold water, are time honored devices which are usually unsuccessful. The object aimed at should be the expansion of the collapsed alveoli, and when the infant makes no respiratory effort this is best accomplished by the insufflation of the lungs with a carbon dioxide-oxygen mixture, carried out by the passage down the trachea of a tube, to the outer end of which a rubber bag filled with the gas mixture is attached. When the breathing is not completely suspended but is weak, and the infant cyanosed, inhalations of a carbon dioxide-oxygen mixture are employed.

PNEUMOTHORAX

Air may enter the pleural cavity through a penetrating wound of the chest (open pneumothorax), as a result of the rupture of an emphysematous vesicle on the surface of the lung or from the extension through the pleura of a lesion of the pulmonary tissue (e.g., tuberculous) or of some other air-containing organ such as the esophagus or stomach. When the intrapleural space contains air but communication with the atmosphere has become occluded the pneumothorax is said to be closed.

OPEN PNEUMOTHORAX

If the opening through the thoracic wall were large and the mediastinum acted as a more or less rigid partition the pleural cavity of the affected side could be considered quite separate from that of the sound side. The lung on the open side would exhibit its elastic properties and recoil, the reserve and residual airs (p. 360) would be expelled and the lung would be in the collapsed state. The opposite lung would be unaffected. Actually, the mediastinum, as a rule, offers little resistance, so

the two pleural cavities, though anatomically separate, act so far as the distribution of pressure is concerned, almost as if they were a single cavity. Consequently when an opening exists in the thoracic wall the yielding mediastinum with its contents—heart and great vessels—moves toward the sound side and the increased intrapleural pressure is transmitted to that side. The negative pressure on both sides of the chest is therefore reduced and both lungs tend toward collapse, but, the pressure on the sound side is always altered less than the pressure on the side of the opening. Graham and his associates found in experiments upon the human cadaver and upon the living dog that an air pressure of +10 cm. of water created in one pleural cavity, caused the pressure in the opposite pleural cavity to rise to between +7 and +8 cm. With the introduction of higher pressure a greater difference was found between the two sides. For example, when a pressure of 48 cm. of water was created in one pleural cavity the pressure on the opposite side was less than 21 cm.

Other things being equal the size of the opening in the chest determines the extent to which the negative pressure becomes reduced. If the chest were immobile and the mediastinum quite unresisting this factor would exert no influence upon the pressure ultimately attained within the thorax. The pressure would finally become atmospheric throughout and complete collapse of both lungs would result. But in the living body the response to pneumothorax is to deepen the respirations. That is, the thorax enlarges and the lungs expand to a greater degree to maintain the negative pressure despite the communication between the chest cavity and the atmosphere. It therefore becomes a matter of competition between the amount of air entering the lungs through the trachea and that entering through the opening in the chest during inspiration. The chest may be compared to a bellows with a hole in its wall, when the bellows is opened (inspiration) the volume of air entering through the leak and that through the nozzle (which is analogous to the trachea) will depend upon the sizes of the respective openings. From this it will be realized that with quite a small opening an intrapleural pressure of practically normal value could be maintained. Even with a very large opening—one exceeding in extent by several inches the cross area of the trachea—though the intrapleural pressure could not be

maintained at its normal value and partial collapse of the lungs would result, nevertheless, a normal volume of air could be drawn in during inspiration. The reason for this is, of course, that the amount of air required for ordinary existence—tidal air—is only a small fraction of the vital capacity. In other words a partially collapsed lung is adequate for ordinary needs. When, however, the opening is so large that the thorax, even when maximally enlarged, cannot expand the lung sufficiently to maintain the tidal air at its normal value, dyspnea and cyanosis will result. It also follows that a patient whose vital capacity is already reduced by disease cannot survive with an opening as large as that which can be tolerated by a person possessing a larger vital capacity. As a result of the work of Graham and his associates the following summary may be made:

(a) Both lungs are affected in a pneumothorax. If the opening is small or if in a closed pneumothorax the pressure is low, the effect upon the two lungs is nearly equal.

(b) A bilateral open pneumothorax is not fatal unless the openings are large.

(c) The size of the opening compatible with life bears a relation to the vital capacity of the subject. Any pneumothorax, unless the opening is very small, would be fatal to a person who before the pneumothorax occurred had possessed a vital capacity little greater than his tidal air. It is therefore possible to perform an intrathoracic operation upon a subject possessing a high vital capacity without distending the lungs by the delivery of air under pressure through the trachea.

(d) The reduction in the intrathoracic "negative" pressure caused by a pneumothorax tends to impede the filling of the right heart and to produce stasis in the venous system (p 168).

(e) Open pneumothorax increases the heat loss of

the body. In dogs the body temperature may fall 2°C half an hour after an opening has been made in the chest. After closure the temperature rises again and within an hour has nearly reached the normal level.

When the mediastinum has been rendered stiff and resistant by previous disease some of the foregoing statements obviously will not apply. Also, if the pleural surfaces on the side of the opening are adherent and so hold the lung out against the thoracic wall a pneumothorax will not result. Or if, as a result of such adhesions the pneumothorax is limited in extent, only a part of the lung is affected.

COLLAPSE OF THE LUNG, ARTIFICIALLY INDUCED
Pulmonary collapse is induced as a therapeutic measure in certain pulmonary lesions, especially tuberculosis in which the disease is mainly confined to one lung. The operations employed are (a) *pneumothorax*, (b) *phrenic avulsion*, (c) *thoracoplasty*. The aim is to render the lung functionless and so to place it at rest. Healing is thereby promoted. Benefits also result from the obliteration of the vessels of the diseased lung. The blood is diverted from the poorly aerated tissue to the healthy lung, whose capillary bed becomes enlarged for the accommodation of the blood. Anoxia and its effects are thereby relieved (see p 426). (a) *Pneumothorax*. The pneumothorax is of the closed type, i.e., air is introduced into the pleural cavity under pressure. The injected air after a time becomes absorbed (see atelectasis, p 431) and the operation must be repeated in order to maintain the lung in the collapsed state. This operation or phrenic avulsion (below) is employed in the comparatively early stages of pulmonary tuberculosis. (b) *Phrenic avulsion*. One or other phrenic nerve is exposed in the neck and sectioned. The lower segment is then seized with forceps and pulled until it gives way, a considerable section of the nerve being removed. The corresponding half of the diaphragm is paralyzed. The latter ascends into the thoracic cavity where it remains fixed at a high level. The capacity of that side of the thorax is reduced and the lung collapsed to a corresponding degree. Mere section of the phrenic in the neck is ineffective as a means of paralyzing the diaphragm since fibers of the 5th cervical nerve (accessory phrenic) join the main trunk of the nerve within the thorax. (c) *Paravertebral thoracoplasty*. This operation is performed in advanced cases with cavity formation and when the pleurae are adherent and consequently an artificial pneumothorax is not feasible. From 1 to 6 inches of the upper nine or ten ribs are removed from the back of the thorax. The chest wall thus rendered plastic sinks inwards and compresses the corresponding lung.

PLEURAL SHOCK Marked slowing of the pulse and a profound fall in blood pressure, which may lead to fatal syncope, sometimes follow puncture of the chest

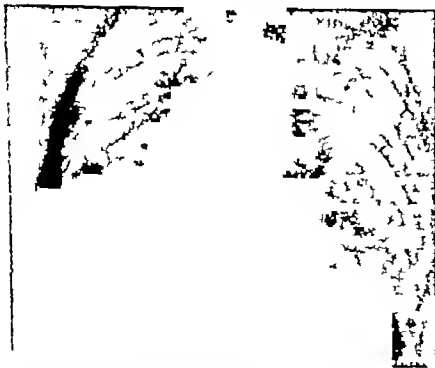


FIG 29.11 Bronchiectasis. X ray photograph after injection of bronchial tree with lipiodol. (After Moll retouched)

wall for the withdrawal of fluid or during the production of pneumothorax. The cardio-vascular reaction is apparently due to a pleural reflex since a similar effect upon heart rate and blood pressure can be induced in animals by stimulating the pleura. It is more commonly seen when the chest is punctured for the withdrawal of fluid than during the injection of air in the production of pneumothorax, a fact among others which argues against its being due to air embolism (footnote, p 173). Injury to the visceral pleura or to the underlying layer of pulmonary tissue appears to be a factor essential for the production of this type of circulatory collapse.

BRONCHIECTASIS

This is the term applied to an abnormal dilatation of the bronchi or bronchioles. The dilatation may be localized or widespread, fusiform, saccular, beaded or uniform in distribution. The contour of the bronchial tubes is readily demonstrated by X-ray after the injection of lipiodol (fig 29 11). Secretion may collect within the dilated lumina and be expectorated at intervals as large quantities of a foul-smelling sputum. This classical sign of the condition is only seen in its advanced stages. More commonly the sputum is small in amount and in the form of yellow "chunky" pieces (Warner).

Causation

Bronchiectasis is practically always secondary to some other affection of the lungs. The primary change leading to the condition is weakening of the bronchial wall as a result of infective processes. The elastic and muscular tissues are atrophic. As we have seen (p 353) the bronchial tree dilates during inspiration and constricts during expiration. The latter movement is, in part at least, an elastic recoil. With the gradual deterioration of the elastic tissue of the bronchial wall, the latter's resiliency becomes progressively less. The lumen

of the tube does not regain its normal caliber after inspiration but remains more or less dilated. According to the most widely held view, further dilation is induced because the pressure within the bronchial lumen during inspiration is higher than the pressure in the surrounding pulmonary tissue and in the pleural space. During expiration, however, the intrabronchial pressure cannot rise above the pressure in the surrounding tissue, the bronchial walls are supported. Coughing or other conditions causing a general rise in intrapleural pressure cannot, therefore, induce bronchiectasis. On the other hand conditions which increase the negative pressure within the thorax, such as the collapse of an area of lung (atelectasis), of a lobe or of entire lung (massive atelectasis), will obviously increase the tendency to bronchiectasis. Incomplete obstruction of a bronchus especially if it exerts a valve-like action, by permitting inspiration but hindering expiration, favors the development of bronchiectasis (see also p 432). The latter is therefore not an uncommon result of the lodgment of a foreign body in a bronchus. On the other hand, it is thought by some (Roberts and Blair) that retention of secretions, as originally suggested by Laennec, combined with inflammatory weakening of the bronchial wall are the important causative factors, and deny the significance of atelectasis and an increase in pleural negative pressure. Whitwell, in an examination of 200 lungs removed at operation, found atelectasis in only 10 per cent.

Fibrosis of a lung or portion of lung is considered by some as a potent cause of bronchial dilatation. It is thought that the fibrosing lung as it shrinks makes equal circumferential traction upon the bronchi or bronchioles and thus leads to their dilatation. It is probable that even in these instances previous weakening of the bronchial wall is the primary fault since bronchiectasis commonly occurs in the absence of fibrosis.

CHAPTER 30

THE AIR OF THE LUNGS

The quantity of air drawn into and expelled from the lungs in quiet respiration is only a fraction ($\frac{1}{4}$ or so) of that which can be inhaled and exhaled during deep breathing. The air which passes in and out of the lungs during ordinary respiration is spoken of as the *tidal air*. It amounts on an average to 500 cc. The average man after he has completed an ordinary expiration, can inhale, by making the deepest inspiration of which he is capable about 3000 cc. This is termed the *complemental air*.¹ If, starting again from a position of rest, i.e., at the end of an ordinary expiration, a forcible expiratory effort is made, about 1000 cc can be expelled. This is called the *supplemental air*. Since in ordinary breathing the lungs must hold this extra quantity of air which can be expelled upon demand it was called the *reserve air* by Hutchinson. The volumes of the complemental and supplemental airs, i.e., the total amount of air which can be exhaled after a maximal inspiration, is called the *vital capacity*. Thus, in round figures,

Tidal air	500 cc
Complemental air (including tidal air)	3000 cc
Supplemental (reserve) air	1000 cc
Vital capacity	4000 cc ²

(see fig. 30 1)

Even after the most strenuous expiratory effort a large quantity of air still remains in the lungs, since collapse of the air cells cannot occur so long as the intrathoracic pressure remains negative (subatmospheric). This is termed the *residual air*. It is present in the lungs after death but is expelled in large part when the pleural cavity is opened and the pressure upon the two sides of the lung becomes equalized. The residual air amounts to from 1000 to 1500 cc. but it shows considerable individual variation.

When the lungs collapse as a result of opening the chest, the small amount of air which still

¹ This of course includes the tidal air. Some take the volume which can be inspired after a normal inspiration as the complemental air. It then does not include the volume of the tidal air.

² The values vary considerably between individuals. Those in the table are for a healthy adult male of average build.

remains entrapped within the air sacs and cannot be expelled by ordinary means is known as the *minimal air*. The minimal air is responsible for the characteristic buoyancy of pulmonary tissue. The lungs of a dead animal float in water and for this reason are popularly known as the "lights". The fact is important in medico-legal practice and is applied as a test to determine whether an infant was still born or had died after having once breathed.

Further partitions of the lung are distinguished by clinicians. Thus *total lung capacity* is the vital capacity plus the residual air, that is, the total volume of air which the lung can hold after a maximal inspiration. It amounts to about 5000 cc. The *minute volume* is the total volume of air breathed per minute, i.e., the volume of the tidal air multiplied by the number of respirations per minute.

The *functional residual air* or *normal capacity* is the amount of air remaining in the lungs at the end of a normal expiration, that is, the supplemental or reserve air plus the residual air. It averages 2500 cc.

THE VITAL CAPACITY (V C) AND ITS VARIATIONS

A relationship between certain body measurements and vital capacity has been shown by several investigators. Hutchinson, a pioneer in the field, invented the spirometer and studied the vital capacity in a large number of individuals, normal and diseased. He demonstrated a relation of vital capacity to height and weight.³

³ Dreyer investigated the vital capacity in a number of healthy individuals and demonstrated a relationship between it and "stem length" (height of body from chair in sitting position) weight, and body surface respectively.

For the average healthy person the relationship between vital capacity and weight is shown in the formula

$$W^{.72}/V C = K$$

When W = weight of body in grams, power $n = 0.72$, $V C$ = vital capacity in cubic centimeters, K = a constant having a value of 0.690. The normal standard for a given weight is calculated as follows

$$W^{.72}/0.690 = V C \text{ in cubic centimeters}$$

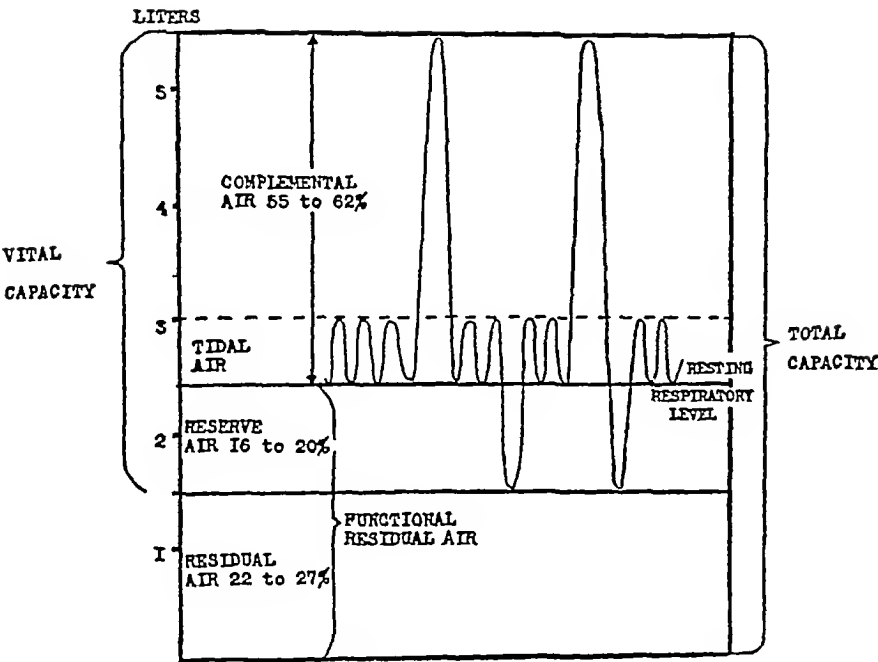


FIG 30 1 Showing subdivisions of lung air (Modified from Christie.)

Pappenheimer and associates (Fed Proc 1950, 9, 602), for the sake of uniformity in nomenclature, have recommended the use of new terms for the subdivisions of the lung volume. These are as follows (the established corresponding terms are in parentheses): *Inspiratory reserve volume* (complemental air, not including the tidal air), *expiratory reserve volume* (supplemental air), *tidal volume* (tidal air), *residual volume* (residual air), *vital capacity* (vital capacity), *inspiratory capacity* (complemental air), *functional residual capacity* (functional residual air).

The most consistent relationship, as shown by Dreyer and by West, exists between vital capacity and surface area (see table below)

A fairly close relationship between total height and vital capacity was also found, the latter expressed as cubic centimeters being 25 times the height in centimeters for men, 20 times for women and 29 times for athletes. For example, an average adult male 170 cm tall would have a vital capacity of $(170 \times 25 =)$ 4250 cc. The surface area of the same individual assuming an average weight for his height, namely, 70 kg, would be 1.80 square meters. Therefore his vital capacity should be $(1.80 \times 2500 =)$ 4500 cc. The difference in the results of the two methods of calculation of the normal standard is therefore around 5 per cent.

The ratios of vital capacity to height and surface area respectively, as found by West, are shown in the following table

	Men	Women	Athletes
Vital capacity, cc per cm height	25.0	20.0	29.0
Vital capacity, cc per square meter of body surface	2500	2000	2800

Occupation as shown by Dreyer and others exerts a potent influence upon the vital capacity, the normal standard of persons employed in sedentary work being considerably lower than

The subject's vital capacity as directly measured by the spirometer is expressed as a percentage of the normal standard for age, sex and occupation

those pursuing more arduous occupations. Dreyer divided his subjects into three classes A, B and C. Class A are those with the maximum vital capacities. Classes B and C have values 90 per cent and 85 per cent respectively of those of Class A. He considers a reduction of 15 per cent below the standard of the class to which the subject belongs as an almost certain indication of some abnormality. It should be mentioned that measurements of chest expansion, as by means of a tape measure may bear little relation to the vital capacity. A subject with powerful muscles can enlarge his thoracic cage to a capacity greater than his lungs are able to fill. His diaphragm instead of taking a full downward stroke is drawn upwards while the viscera are forced into the extra-thoracic space by a strong contraction of the abdominal muscles.

The vital capacity is reduced in many diseased conditions especially those involving the respiratory and cardio-vascular systems. Among these are

- (a) Conditions which directly involve the lungs, e.g., pneumonia, pulmonary tuberculosis, emphysema, etc.
- (b) Mechanical interference with the enlargement of the thoracic cavity, e.g., as by abdominal conditions impeding the movements of the diaphragm or abnormalities of the thoracic walls. The movements of the thoracic walls may be seriously restricted by abdominal or pleuritic pain.
- (c) Intrathoracic conditions which encroach upon

the space normally occupied by the lungs, e.g., pericardial or pleuritic effusions, pneumothorax, tumors, etc

(d) Heart disease Engorgement of the pulmonary vessels causes an encroachment upon the alveolar spaces and so reduces their capacity. Pulmonary edema involving especially the lung bases is a contributory factor in the reduction of the vital capacity in cardiac cases. Peabody found that when in heart disease the vital capacity reached 40 per cent of the normal, the subjects were almost constantly dyspneic or showed dyspnea upon the slightest exertion (p. 415).

The vital capacity, the functional residual air and the residual air and consequently the total lung capacity are reduced in the recumbent posture.

METHODS OF MEASUREMENT

It must be emphasized that any reduction of the vital capacity of an individual is of much greater significance than an apparently abnormal value encountered at the first examination. The vital capacity, the complemental air, or the reserve air can be readily measured in a simple spirometer. For the determination of residual air or the volume of the lung at any phase of respiration the modification of Humphrey Davy's dilution method described by Van Slyke and Binger is recommended. Oxygen to which a known volume of hydrogen has been added is breathed in and out from a spirometer. The carbon dioxide is absorbed by sodium hydroxide. A sample of the mixture is analysed and the ratio of N_2 to H_2 determined, since,

$$(1) \quad \frac{N_2 \text{ in sample}}{H_2 \text{ in sample}} = \frac{\text{Total volume } N_2}{\text{Total volume } H_2}$$

and (2) the total volume of H_2 being known, the total volume of N_2 in the lungs can be calculated, then in as much as nitrogen forms 79 per cent of the lung air, $(\text{vol } N_2 / 79) \times 100 = \text{lung volume}$

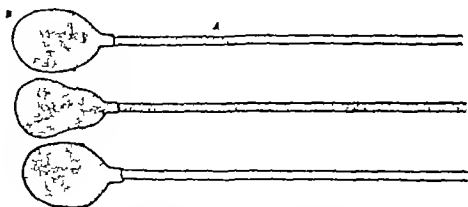


FIG. 302 Description in text.

ALVEOLAR AIR

This is the air contained in the air sacs and alveoli. Its gases come into equilibrium with those in the blood of the pulmonary capillaries. The alveolar air is being continually renewed through the to and fro effect of the respiratory movements. But since the lungs are not emptied during expiration it is quite clear that complete renewal of the air cannot be effected by a single respiration. The ventilation of the lungs occurs by the mechanical mixing of the inspired air with the lung air and by the slower process of diffusion. During expiration the air of the upper respiratory passages is first expelled, then follows a part of the air of the bronchial tree. At the end of expiration the air which has been in the distended alveoli, since these have been reduced in capacity, now overflows into the bronchioles and bronchi. At the next inspiration this air is swept back again into the air sacs, its place being taken by fresh atmospheric air. The latter as compared with the alveolar air has a high oxygen pressure and a low pressure of carbon dioxide. Oxygen therefore diffuses from the inspired to the alveolar air and carbon dioxide from the alveolar air to the inspired air. The mechanical mixing of the inspired air with the air already in the lungs occurs, chiefly, at the beginning of inspiration and the end of expiration. Mixing is also brought about through the convection currents set up by the entrance of the cooler atmospheric air.

The process by which the air in the alveoli is renewed may be made clear by the illustration in figure 302. A is a glass tube to which is attached a compressible bulb B. If the bulb is filled with smoke and then compressed, the smoke will not be completely expelled by a single compression. A portion only will issue from the end of the tubing. The rest will fill the compressed bulb and tubing. When the bulb is released again smoke will be drawn back into the cavity and the tubing will be filled with fresh air. The smoke in the bulb will be diluted through mechanical mixing and diffusion. After a series of compressions the smoke would be completely cleared away. If the bulb were supplied steadily with smoke through the bulb end and rhythmical compressions and decompressions of the bulb maintained at a definite rate, a certain constant dilution of the smoke, but not complete clearance, would result. The smoke entering the bulb may be taken to represent the carbon dioxide passing from the blood into the alveolar spaces. Henderson and his associates have shown that when one gas displaces another from a cylindrical tube the interface between

the two is not a plane surface. On the contrary, the displacing gas advances in the form of a cone or spike leaving a layer of the gas which is being displaced clinging to the tube's walls. This fact may be demonstrated by displacing the air from a glass tube by a puff of smoke (fig 30 3). When the current of smoke ceases mixing occurs.

Obtaining a sample of alveolar air

It is clear from the foregoing illustration that in order to obtain a sample of pure smoke as it issues from the end of the tubing all the air must first have been swept out. Similarly in order to obtain pure alveolar air a forced expiration must be made and a sample collected as close as possible to the subject's mouth. This is accomplished through the use of the Haldane tube, which consists of a length of hose-pipe 1 inch in diameter and 3 feet long and provided with a mouth piece and sampling bulb as shown in figure 30 4. The subject after a normal expiration makes a forced breath down the tube. Toward the end of this act the sampling bulb, which has been previously evacuated, is opened for a moment in order to permit the expired air to be drawn in, and then closed. An all glass air-tight syringe may be used instead of the sampling tube. The exact technique used in research is somewhat more complicated than that described above.

The following are the average compositions of dry inspired, expired and alveolar airs reduced to standard temperature and pressure (S T P)

	Volumes per cent		
	Inspired (atmospheric) air	Expired air	Alveolar air
Oxygen	20.94	16.3	14.2
Carbon dioxide	0.04	4.0	5.5
Nitrogen (including argon 0.94 per cent and other rare gases)	79.02	79.7	80.3

The atmospheric air contains an average of less than 1 per cent water vapor. The lung air contains about 6.2 per cent.

THE DEAD SPACE AIR

The respiratory passages extending from the nostrils to and including the terminal bronchioles (p 344) constitute what is known as the *anatomical dead space*. The walls of these passages are relatively thick and no gaseous interchange between blood and air can occur across them. The capacity

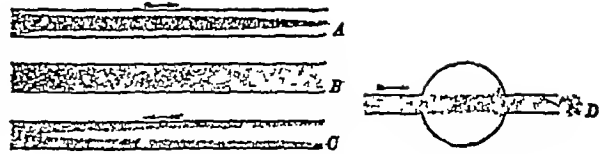


FIG 30 3 A shows a 'spike' of smoke moving through a glass tube. B shows the condition when the current is suddenly stopped and mixing occurs instantaneously. C shows clear air drawn in through a glass tube. D shows how a column of smoke crosses a bulb with little mixing or sweeping out of the air within it. (After Yandell Henderson and associates)

of the anatomical dead space, which was first estimated by Loewy from a cast of the respiratory passages of a dead subject, is about 150 cc and though variable between individuals is relatively constant for the same person. During an ordinary *inspiration* about 500 cc of air (tidal air) are inhaled. A proportion of this fresh air (150 cc) is used to fill the dead space, the air previously in the dead space being, as mentioned above, drawn into the alveoli together with the remainder (350 cc) of the inspired air. This 350 cc is added to a large volume of air (residual and supplemental airs amounting to about 2000 cc) which remains in the lungs after the previous expiration. During *expiration* alveolar air high in carbon dioxide and low in oxygen is forced from the air sacs. Part of this is exhaled to the outside and part remains within the dead space from which it has displaced the fresh air of the previous inspiration. The expired air is therefore a mixture of alveolar and inspired airs.

The physiological dead space and its estimation

The anatomical dead space as described in the preceding section varies relatively little in capacity though a moderate increase or decrease may occur as a result of bronchiolar dilatation or constriction. The *physiological*, *virtual* or *effective dead space* are terms applied to the total space within the lung which just prior to expiration contains perfectly fresh air, that is, air which has not diluted the alveolar air or come into contact with the respiratory epithelium. The volume of the physiological dead space varies with the depth



FIG 30 4 Haldane tube and sampling bulb

RESPIRATION

of respiration. In shallow breathing fresh air may not be drawn as far as the terminal bronchioles. The physiological dead space would then be less than the anatomical. In ordinary breathing the physiological dead space is considerably larger than the anatomical (i.e., fresh air penetrates beyond the terminal bronchioles), and in deep breathing is greatly increased as a result of the passive dilation of the respiratory bronchioles and alveolar ducts which after a deep inspiration receive fresh atmospheric air. The advancing margin of the latter, however, as a result of its spike form, which has been referred to above, is prevented from coming into direct contact with the respiratory epithelium by a layer of vitiated (alveolar) air. Its gases, therefore, like those of the air of the anatomical dead space can reach the respiratory epithelium only by diffusion which, as compared with mechanical mixing, is a slow process.

Since the physiological dead space is really a measure of the depth to which fresh (inspired) air penetrates and remains unmixd with alveolar air, its volume can only be determined by calculation. It will be equal to the volume of expired air less that of the alveolar air in the expiration. The volume of the alveolar air is

derived from the volume of the expired air and the ratio of the percentages of CO_2 in alveolar and expired airs. For example, let us say that (a) the volume of air expired in a single breath is 500 cc and (b) expired air contains 4 per cent carbon dioxide and (c) alveolar air contains 6 per cent carbon dioxide. Then the volume of the alveolar air in the expiration is $(500 \times \frac{4}{6}) = 333$ cc. The volume of the dead space therefore must be $(500 - 333 =) 167$ cc. During deep breathing the ratio of the percentage of carbon dioxide in the alveolar air to that in the mixed expired air is about the same as during quiet breathing. Since the volume of the expired air is increased, the effective dead space as calculated by the above method would therefore be increased also.

Example

Volume of expired air in a single breath	2000 cc.
Carbon dioxide in expired air	4 70 per cent
Carbon dioxide in alveolar air	6 20 per cent
Volume of alveolar air is therefore	$2000 \times (\frac{4.70}{6.20})$
$= 1516$ cc and volume of effective dead space $=$	
$2000 - 1516 = 484$ cc	

CHAPTER 31

THE PHYSICAL PRINCIPLES GOVERNING THE RESPIRATORY EXCHANGES

The interchange of gases which occurs between the atmosphere and the air in the lungs and between the latter and the pulmonary capillaries is termed *external respiration*. Carbon dioxide diffuses from the blood to the air in the air sacs which in turn is freshened by inspired air (p 367). Oxygen diffuses from the inspired air to the alveolar air and from this into the blood. The gaseous exchange between the systemic blood and the tissues is reversed—oxygen passing from the blood to the tissues and carbon dioxide from the tissues to the blood. This is called *internal respiration*. The gaseous exchanges are governed by physical laws—upon differences in pressures (tension) between the blood plasma on the one hand and the air or tissue fluids on the other.

Hemoglobin serves for the storage and transport of oxygen, sodium bicarbonate serves corresponding purposes for carbon dioxide (see p 395). The actual interchange of these gases depends, however, upon their relative tensions in the different media, i.e., in the lung air, and in simple solution in the plasma and tissue fluids.

THE KINETIC THEORY OF GASES—DIFFUSION, PARTIAL PRESSURES, ABSORPTION COEFFICIENTS

Any quantity of a gas when placed in a container of whatever size expands its volume until limited by the boundaries of the confining vessel. This fact is explained upon the theory that the gas molecules are in continuous motion, moving through space at high velocity and being deflected from their course only upon coming into collision with other gas molecules or with the boundaries of the space itself which they strike and from which they rebound repeatedly. Such movements constitute a bombardment upon the confining walls which is responsible for what is called the pressure or tension of the gas. The greater the number of molecules in any given space the more frequent, obviously, will be the bombardments. So, if the capacity of the space is reduced the molecules are brought closer together. The rate of bombardment upon a unit of surface will increase and the pressure of the gas will rise (see *Boyle's Law* below). A

rise in temperature increases the velocity of the molecular movements, increases the rate of bombardment and the force of the impacts. The pressure in consequence increases (*Charles' Law*).

In all circumstances, the gas molecules as a result of their movements will in time distribute themselves evenly throughout the space in which they are confined and consequently the pressure will be the same upon all parts of the limiting surface. In other words, although at the start the molecules may be in greater numbers (concentration) in one part of the space than in another, even distribution is soon brought about and the pressure throughout all parts of the space becomes equal. This behavior of a gas whereby the equalization of its molecular concentration occurs is spoken of as *diffusion*. If we should deal with a mixture of two or more different gases instead of with a single gas it would be found that each component in the mixture behaved as though present alone. Its molecules would become distributed evenly throughout the mixture and its pressure would depend upon its concentration without regard to the concentrations of the other component gases (*Law of partial pressures*).

If two samples of a gas, of different concentrations, be placed one on each side of a membrane permeable to that gas, diffusion also occurs until the tensions on the two sides of the membrane are equal, as in the case of a gas mixture each gas behaves as though present alone.

When a gas or a mixture of gases lies in contact with the surface of a liquid, the molecules of each gas penetrate the liquid and become dissolved in it until the tensions of that particular gas within and without the liquid are equal. The gas is then said to be equilibrated with the liquid (*Law of solubility of gases*). On the other hand, if the liquid be now exposed to a lower pressure of the gas the molecules which had undergone solution at the higher pressure escape until equilibrium is restored at the lower level. Soda water, for example, is water which has been equilibrated with carbon dioxide at a high pressure. When the cork is removed from the bottle containing the surcharged liquid, effervescence occurs. Mole-

cules of carbon dioxide are given off until the pressure of the dissolved gas equals that of the carbon dioxide (0.3 mm Hg) in the atmosphere. A gas at different tensions in two liquids also comes into equilibrium whether the liquids are in direct contact or are separated by a membrane permeable to the gas. Also the actual amount of gas which will undergo solution at a given pressure varies with the particular gas and with the liquid. If distilled water for example be exposed to oxygen at a pressure of 760 mm and a temperature of 0°C each 100 cc. will take up 4.9 cc. of the gas. Oil, on the other hand, under the same conditions of pressure and temperature will absorb a great deal more. Therefore if samples of water and oil are exposed to the atmosphere though the gas pressures in the three media are identical, the volumes of the atmospheric gases in 100 cc. (volumes per cent) of each medium will be widely different.

The quantity of a gas (measured at standard temperature and pressure) which can be absorbed by 1 cc. of a liquid at 760 mm Hg is called the *absorption coefficient* of the gas for that particular liquid. Thus the absorption coefficient of oxygen in water at 0°C is 0.049 and of carbon dioxide 1.71, the coefficient varies inversely with the temperature (see table 26). The presence of dissolved solid substances in the water will reduce the absorption coefficient of these gases. The values for the body fluids are therefore slightly less than those given above. Thus the coefficient of absorption of oxygen in plasma at body temperature (37°C) is 0.024 and of carbon dioxide 0.510.

The rate of diffusion of a gas through a liquid is in direct proportion to the absorption coefficient of the gas in that liquid and inversely proportional to the square root of its molecular weight. The diffusion rate of carbon dioxide through a wet membrane is about 30 times greater than that of oxygen when the two gases are under identical conditions. For general physiological work the *diffusion coefficient* of oxygen has been defined by

Krogh as the number of cubic centimeters of the gas which will diffuse 0.001 mm distance over a square centimeter of surface, per minute, at a pressure of 1 atmosphere. It varies for different tissues and body fluids and increases 1 per cent per degree Centigrade.

The diffusion coefficients for oxygen through the following materials at body temperature were found

Water	0.51
Gelatin 15 per cent	0.45
Muscle	0.31
Connective tissue	0.18

The diffusion coefficient of oxygen through the pulmonary epithelium of the intact animal is defined as the number of cubic centimeters which are absorbed per minute per millimeter of mean pressure difference between the blood and alveolar air. The coefficient varies in different individuals from 23 to 45 during rest and from 37 to 56 during exercise. If, for example, the mean difference between the oxygen tensions of the blood and alveolar air should be 10 mm Hg and the coefficient 25 mm, then $(10 \times 25 =) 250$ cc. of oxygen will be absorbed per minute. The increase in the coefficient during exercise is ascribed to the opening up of more capillaries. The differences observed between individuals are probably dependent upon the sizes of the alveoli, the thickness of the alveolar epithelium and the mean capacity of the lungs. The diffusion coefficient of carbon dioxide through the alveolar epithelium is around 500 during rest and 800 during exercise.

SUMMARY OF THE GAS LAWS

(1) *Boyle's Law*. When the volume of a gas is altered, the temperature remaining constant, the pressure varies inversely, i.e., the product of the pressure and the volume remains constant. If the space wherein a certain gas is confined be reduced by half, the gas pressure is doubled and vice versa.

(2) *Law of Charles (or Gay Lussac)*. For each rise in temperature of 1°C a gas kept at constant pressure expands by $\frac{1}{273}$ of its volume at 0°C. The volume of a gas at constant pressure is therefore proportional to its absolute temperature (-273°C).

(3) *The Law of Partial Pressure (Dalton's Law)*. The pressure exerted by a gas in a mixture of gases is equal to the pressure which the same quantity of that gas would exert were no other gases present. It follows that the total pressure of a mixture of gases is equal to the sum of the pressures of its component gases. For example, the atmosphere (dry) exerts a pressure of 760 mm Hg. The gases of which it is composed—oxygen, nitrogen and carbon dioxide are present in the proportions of 20.96 per cent, 79 per cent and 0.04 per cent respectively. The partial pressure ex-

TABLE 26

Absorption coefficients of various gases in distilled water at different temperatures

TEMPERATURE	OXYGEN	CARBON DIOXIDE	CARBON MONOXIDE	NITROGEN
0	0.049	1.71	0.035	0.024
20	0.031	0.87	0.023	0.016
40	0.023	0.53	0.018	0.012

erted by oxygen is therefore $(20.96/100) \times 760 = 159.2$ mm Hg and of carbon dioxide $(0.04/100) \times 760 = 0.30$ mm Hg

Air in contact with water is continually receiving water molecules from the surface of the liquid. This water vapor follows Dalton's Law exerting a pressure independently of the other gases, and proportional to the quantity present in the air. The higher the temperature the greater is the quantity of water which the air will hold before becoming saturated and the greater consequently will be the tension of aqueous vapor.

The air in the lungs has a temperature of about 37°C and is usually stated to be fully saturated with water vapor, the latter, therefore, exerts a pressure of 47 mm Hg.¹ The air after leaving the lungs falls in temperature, some of the water vapor condenses and the latter in consequence is much less. The tension of water vapor in room air (18°C) would be no more than 15.5 mm Hg, even though the air were fully saturated, and is usually around 4 or 5 mm Hg. The aqueous tensions of air (saturated) at various temperatures are given in table 27.

In the measurement of the respiratory gases the volumes are expressed dry (i.e., less the aqueous vapor, though actually no correction for this is required) and at standard temperature and pressure (STP)—760 mm Hg and 0°C . The individual gases, carbon dioxide or oxygen, are then expressed as percentages of this dry volume.

In order to arrive at the tension of one or other gas from its percentage in dry air, the figure for the barometric pressure less the aqueous tension must of course be used as the basis for calculation. For example, if the carbon dioxide percentage in a sample of alveolar air (dry) is 5.6 per cent and the barometric pressure, (and so of course the total gas pressure of the alveolar air) is 760 mm Hg, then the tension of carbon dioxide in the alveolar air must be

$$(5.6/100) \times (760 - 47) = 39.9 \text{ mm Hg}$$

Similarly when the O_2 percentage in dry alveolar air is 14.2 the oxygen tension is

$$(14.2/100) \times (760 - 47) = 101.2 \text{ mm Hg}$$

(4) *Henry's Law of the Solution of Gases* If the temperature remains constant then the quantity of a gas which goes into solution in any given liquid is proportional to the partial pressure of the gas.

¹ Christie and Loomis from direct measurements have obtained a lower value for the aqueous tension of alveolar air than the usually accepted one of 47 mm Hg, (namely 45 mm Hg). They claim that the alveolar air is not fully saturated and that the temperature of the lung is lower than has been assumed. Hyperpnea they found reduced the aqueous tension by as much as 7 mm Hg. Holding the breath increased it by 0.5 mm Hg.

TABLE 27

Tension of aqueous vapor and water in grams in moisture-saturated air at different temperatures

TEMPERATURE	TENSION OF AQUEOUS VAPOR	WATER PER CUBIC METER OF AIR
$^{\circ}\text{C}$	mm Hg	grams
0	4.6	4.9
5	6.5	6.8
10	9.1	9.4
15	12.7	12.8
20	17.4	17.2
30	31.6	30.1
37	47.1	

TABLE 28

BAROMETER 760 MM Hg	PARTIAL PRESSURE ²		
Gas	Inspired air	Expired air	Alveolar air
	mm Hg	mm Hg	mm Hg
Oxygen	158.25	116	100
Carbon dioxide	0.30	28	40
Nitrogen	596.45	569	573
Aqueous vapor	5.00	47	47
Totals	760.00	760	760

THE EXCHANGE OF GASES IN THE LUNGS

THE PARTIAL PRESSURES OF THE GASES IN THE LUNG AIR

In table 28 are given average figures for the partial pressures of oxygen, carbon dioxide and nitrogen in inspired, expired and alveolar airs. The fall in oxygen pressure from inspired to alveolar air and in the reverse direction for carbon dioxide will quite evidently promote the free interchange of these gases across the pulmonary epithelium. The interchange of gases between inspired and alveolar airs is reflected in the intermediate values shown for the gas pressures in the expired air.

The alveolar oxygen and carbon dioxide tensions tend to vary with the minute volume (p. 360). During voluntary hyperpnea the CO_2 tension

² These are general figures. There is considerable variation between individuals and under different conditions. The range of gas pressures in the alveolar air for man at rest is from 97 to 108 mm Hg of oxygen, and from 35 to 45 mm Hg for carbon dioxide.

Minute amounts of argon, helium, neon and other rare gases, are present in the atmosphere and are included in the figure given for nitrogen. They are inert in so far as respiration is concerned, nor do they appear to be essential for any physiological process.

TABLE 29
Gas tensions in arterial and in venous blood

	TENSION	
	Venous blood	Arterial blood
	mm Hg	mm Hg
Oxygen	37	100 ²
Carbon dioxide	46	40
Nitrogen	573	573
Water vapor	47	47
Totals	703	760

falls and the O₂ tension rises. When the breath is held or during periods of apnea changes of a reverse order occur.

Nitrogen, so far as respiration is concerned is inert. A small but constant amount (about 0.83 volume per cent) of the gas is taken up and dissolved in the plasma but it is neither used nor produced within the body, the quantities in arterial and in venous blood being identical. It will be noted, however, that the percentage of this gas is higher in alveolar and expired air than in inspired air (p. 363). This is due not to any absolute increase in the quantity of nitrogen but to the reduction of the total volume of the respiratory gases resulting from the greater quantity of oxygen absorbed than of carbon dioxide put out.

THE PARTIAL PRESSURES OF OXYGEN AND CARBON DIOXIDE IN BLOOD

In table 29 are given the tensions of oxygen and carbon dioxide in arterial and in venous blood. The venous blood, it will be seen, has a lower tension of oxygen (by 60 mm.) than the alveolar air but a higher tension of carbon dioxide (by 6 mm.). It is to be remembered that the pulmonary capillaries and the air in the alveoli are separated by delicate membranes freely permeable to these gases. The pressure gradients are favorable to a rapid inward diffusion of oxygen (from alveolar air to blood) and an outward diffusion of carbon dioxide (from blood to alveolar air). Equilibrium is quickly established between the respiratory gases in the alveolar air and in the blood of the pulmonary capillaries. Since the diffusion coefficient of carbon dioxide through the

pulmonary membrane is much higher than that of oxygen, the pressure gradient of the former gas need not be so high for this to occur. It will be seen from tables 28 and 29 that the tension of carbon dioxide in the blood leaving the lungs, i.e., in the mixed arterial blood, is the same as that in the alveolar air. Indeed, it is the usual practice, when one wishes to know the arterial CO₂ tension, to determine that of the alveolar air and assume that the two are identical. It has been shown, however, by Bock and associates that a slight difference (0.5 mm. Hg) does exist.

The oxygen tension of the mixed arterial blood returned from the lungs of man is only slightly lower than that of the alveolar air. Comroe and Dripps, in a series of normal human subjects, obtained a value of 97.4 Hg for the alveolar air and 97.1 Hg for the arterial blood⁴—an alveolar-arterial difference of only 0.3 Hg. Comparable values for man were obtained by Lilienthal and associates.

Gaseous equilibrium is attained not instantaneously but progressively along the course of the pulmonary capillary. Diffusion is rapid at the venous end but as blood and alveolar air approach equilibrium the diffusion process necessarily becomes slower. The length of the capillary and the rate of the blood flow through it are factors which must determine the extent to which equilibrium occurs. Roughton obtained a figure of 0.75 seconds for the time of passage of the blood through the pulmonary capillaries for normal men at rest, this was shortened to 0.34 seconds during heavy work. The total volume of blood in the pulmonary capillaries at any moment is also, owing to their distension, much greater than during rest (p. 332), during arduous exercise the quantity of blood in the lungs may be double that during rest.

The tensions of oxygen and carbon dioxide in human arterial blood may be determined by bringing a small bubble of air into contact with a sample of the blood obtained by arterial puncture. The method employed for the purpose is an adaptation to man (Barcroft and Nagahashi) of a method originally invented by Krogh for animals (see fig. 311). After the gases in the air have come into equilibrium with those dissolved in the blood the small bubble is analyzed and its

² The range of arterial gas tensions for man at rest is from 93 to 102 mm. Hg for oxygen, and from 35 to 45 mm. Hg for carbon dioxide.

⁴ Since virtually no oxygen is released from the blood in the arterial system, a sample taken from an artery will have the same O₂ tension as that in a pulmonary vein.

composition determined. The gas tensions are then calculated from their percentages (p 367). Since no gaseous exchanges occur across the walls of the arteries, the tensions so determined are those of the mixed arterial blood of the pulmonary capillaries. It has already been mentioned that the arterial carbon dioxide tension is usually obtained from an analysis of alveolar air. The gas tensions of the blood coming to the lungs—the *mixed venous blood*—may be determined in man by direct or indirect methods (ch 26). In animals mixed venous blood may be obtained from the right ventricle by puncturing the chest wall with a hollow needle attached to a syringe.

THE VOLUMES OF OXYGEN AND CARBON DIOXIDE IN BLOOD

Knowing the respective absorption coefficients for oxygen and carbon dioxide in plasma, and the gas pressures, the volume of each gas in *simple solution* in 100 cc of plasma can be readily calculated (table 30). For example, the absorption coefficient of oxygen in plasma is 0.023 at body temperature and at a pressure of 760 mm Hg. At the partial pressure of oxygen in arterial blood the plasma should hold in solution $\frac{100}{760} \times 0.023 = 0.003$ cc of oxygen per cubic centimeter or 0.3 volume per cent. The absorption coefficient of carbon dioxide in plasma (0.51) or of whole blood (0.48) is higher than that of oxygen but the partial pressure of carbon dioxide to which arterial blood is exposed in the lung is lower (40 mm Hg) than that of oxygen. The quantity of the former gas in simple solution in whole blood is therefore $\frac{40}{760} \times 0.48 = 2.5$ volumes per cent.

About 19.5 volumes per cent of oxygen and from 40 to 50 volumes per cent of carbon dioxide can be removed from arterial blood. The proportions of these gases present in simple solution must, therefore, be only a small fraction of the quantities held in the blood in other ways. They are present in chemical combination—oxygen with hemoglobin and carbon dioxide mainly as bicarbonate (see chap 33).

Blood normally contains about 15 grams of hemoglobin per 100 cc. Since 1 gram of hemoglobin carries a maximum of 1.34 cc. of oxygen, arterial blood would, therefore, when saturated to its full capacity, contain about 20 cc. of oxygen. Blood as it leaves the lungs is about 97.5 per cent

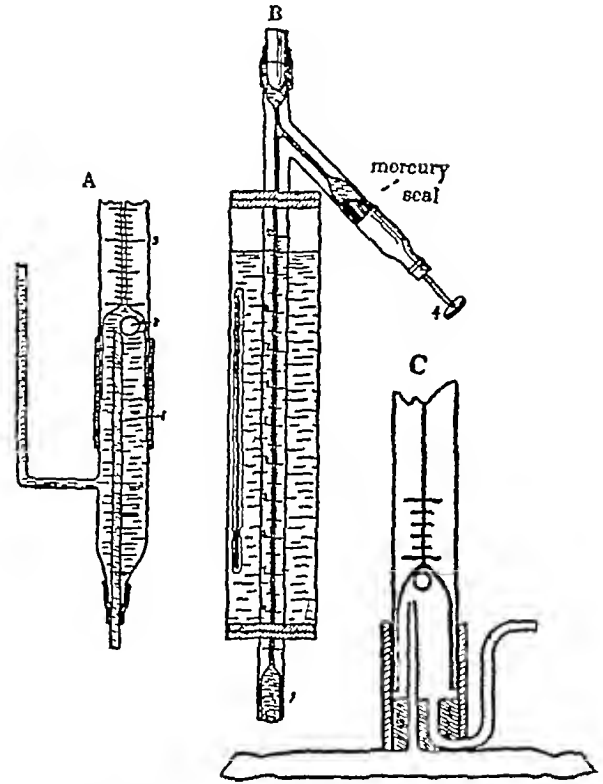


FIG 31.1 Krogh's microtonometer. A is an enlarged view of an equilibration chamber constituting the lower part of B(1) but which is shown only in part in the latter drawing. A is filled with saline solution through the side tube 1, a narrow tube into which the specimen of blood is introduced, 2, an air bubble. The lower end of 1 is connected by rubber tubing to a cannula inserted into a blood vessel, the blood issues from a narrow opening at the upper end of the tube 1 in a fine jet and plays upon the bubble. After the gases in blood and bubble have come into equilibrium the latter is drawn into the graduated capillary tube shown in B, by means of the screw-plunger (4), and analyzed according to the following procedure. The volume of the bubble is first measured, it is next drawn into a solution of KOH which absorbs the CO₂, then returned to the graduated tube and measured again, the difference between the two measurements gives the amount of CO₂ which it contained. The bubble is then passed through a solution of potassium pyrogallate, to absorb the O₂, and its volume measured a third time. C is a model with an attached cannula which can be inserted directly into a blood vessel.

TABLE 30

Volumes per cent (cubic centimeters of gas per 100 cc blood) of oxygen and carbon dioxide in arterial and in venous blood

GAS	VENOUS BLOOD		ARTERIAL BLOOD	
	Total	In simple solution	Total	In simple solution
Oxygen	12-14	0.1	19.5	0.3
Carbon dioxide	58	3.0	48	2.5

saturated with oxygen so that the hemoglobin holds at this degree of saturation about 19.5 cc of oxygen per 100 cc. of blood⁵

The photo-electrical determination of the oxygen saturation. In man, the oxygen saturation can be determined in the blood as it circulates through a translucent part such as the pinna of the ear, by means of the *oximeter*. This instrument, originally suggested by the work of Squires, has been developed by Goldie and by Milikin. The method is based upon the fact that red light (wave length 600 to 7500 millimicrons) is transmitted readily by oxyhemoglobin, but only slightly by reduced hemoglobin. The apparatus consists of an electric light bulb and a photo-electric cell placed, respectively, at the front and back of the pinna. The heat of the bulb dilates the vessels so that the oxygen saturation of the blood in the capillaries becomes that of the arterial blood, the light transilluminates the tissue and, after passing through suitable filters, is received by the photo-electric cell. The electric current set up is measured galvanometrically, and the actual

per cent of oxygen saturation is calculated by means of the calibrations of the instrument with values established by the Van Slyke method. This method has proved of great value in determining the oxygen saturation of the blood drawn by catheter from the right auricle, ventricle and pulmonary artery in the diagnosis of congenital cardiac defects (p. 433). Groome and his associates have devised a more convenient apparatus for this purpose. The blood is transilluminated by passing it through a tube composed of polythene attached to the catheter, on the opposite side the transmitted light is registered by photo-electric cells, placed, respectively, behind red and near infra red (wave length 750-900 millimicrons) filters. The oxyhemoglobin concentration is calculated from the value registered by the first cell, the total hemoglobin from that given by the second cell. From the ratio of these values the oxygen saturation is estimated.

The quantity of oxygen or of carbon dioxide contained in a given sample of blood (*oxygen or carbon dioxide content*) is determined by transferring the sample to a blood-gas apparatus (Haldane or Van Slyke) and then freeing all the oxygen from the hemoglobin by the addition of potassium ferricyanide, or all carbon dioxide from combination by the addition of acid. Precaution must be taken not to permit the sample to come in contact with air. Since the cell wall is not freely permeable to $K_2Fe(CN)_6$, the corpuscles should be first laked by ammonia or saponin solution. Haldane showed that all the oxygen in blood is liberated by this procedure and the oxyhemoglobin is turned into methemoglobin (p. 58). The *oxygen capacity* of a sample of blood is calculated by exposing it to air or oxygen and determining the amount which it then contains, that is, when the hemoglobin is completely saturated. The ratio of oxygen content to oxygen capacity $\times 100$ gives the percentage saturation of the blood with oxygen. Since the quantity of oxygen (1.34 cc.) which will combine with 1 gram of hemoglobin is known, the hemoglobin content of a specimen of blood may be calculated from the quantity of O_2 in the blood when fully saturated, i.e., from its oxygen capacity. For example, if a sample of blood has an oxygen capacity of 10 volumes per cent its hemoglobin content is $(10/1.34 =) 7.5$ grams per cent or about 50 per cent of the normal.

⁵ Since the blood in passing through the capillaries of an air sac comes into equilibrium with the air of that air sac as described on page 310 the statement that the arterial blood has a tension less than that of the alveolar air and is only from 94 to 96 per cent saturated, requires explanation. The discrepancy is explained as follows. The air sacs are not all ventilated to the same extent, in some the O_2 tension is higher, in some lower than that of the alveolar air as determined upon a sample. In other words the O_2 tension of a sample of alveolar air is an average of the O_2 tensions of the air in all the air sacs. But when we come to consider the blood coming, not from a single air sac, but from the lung as a whole—i.e., the *mixed arterial blood*—it becomes evident that while an under ventilated air sac will lower the oxygen tension and so the oxygen saturation of the hemoglobin, an over ventilated one cannot compensate this effect to any significant extent. The O_2 in solution (upon which the O_2 tension directly depends) is in equilibrium with that combined in hemoglobin and it is evident from the shape of the O_2 dissociation curve of hemoglobin that a rise in oxygen tension above 100 mm Hg will saturate the blood very little more whereas a fall of 20 mm Hg or so will reduce the saturation very materially.

The variation between individuals in the O_2 saturation of the arterial blood is attributed to the different degrees to which uneven ventilation of the air sacs occurs and also to slight differences in the shapes of the dissociation curves. It was first shown by Barcroft that the O_2 dissociation curve is not precisely the same for all persons. The dissimilarities are due apparently to slight differences in the chemical constitution of the globin part of the hemoglobin molecule.

CHAPTER 32

THE TRANSPORT AND DELIVERY OF OXYGEN TO THE TISSUES, INTRACELLULAR OXIDATIONS AND ENERGY TRANSFER, THE CARRIAGE OF CARBON DIOXIDE

Hemoglobin serves for the storage and transport of oxygen. The small amount of oxygen in simple solution (about 1 per cent of the total) is negligible when one considers the oxygen requirements of the tissues even during rest (250 cc per minute). If the blood could hold no more than this it would be necessary for some 120 liters to circulate through the tissues each minute, even assuming that all the oxygen were given up during each circulation. Nevertheless, the gas in simple solution is of the utmost importance since it is in equilibrium with the alveolar air on the one hand and on the other determines the quantity of oxygen which shall be held in combination with the hemoglobin. This will be made clear from a study of the oxygen dissociation curve for hemoglobin.

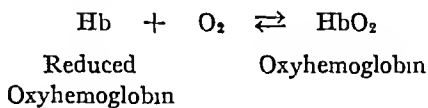
THE DISSOCIATION CURVE OF HEMOGLOBIN

Oxygen enters into chemical combination with the iron of the hemoglobin molecule (see p 57) to form *oxyhemoglobin*. Each atom of the metal unites with two atoms of oxygen. The remarkable feature of the union of oxygen with hemoglobin is the readiness with which the gas is released from combination when its tension in the surrounding medium is reduced. Hemoglobin from which the oxygen has dissociated is called *reduced hemoglobin*. The relationship between the partial pressure of oxygen and the percentage saturation of the hemoglobin with the gas—i.e., the proportion of oxyhemoglobin to reduced hemoglobin—can be shown in the form of a curve—the *oxygen dissociation curve of hemoglobin*. The curve for a solution of hemoglobin in distilled water is obtained in the following manner. Several samples of the hemoglobin solution are placed each in a separate closed vessel known as a tonometer (fig 32 1). The respective samples are then exposed to known oxygen tensions 0, 10, 20, 40 and 100 mm Hg. The tonometers are rotated continuously in a water bath at body temperature. The solution is thus spread out as a thin film over the interior surface of the vessel. After equilibrium has been attained, the proportion of oxy- to reduced hemoglobin is

determined. When 100 per cent saturated—that is, when the hemoglobin has taken up oxygen to its full capacity—the solution contains about 20 volumes per cent of oxygen. When a quarter or half saturated it therefore contains 5 or 10 volumes respectively. The results are plotted on a chart with the oxygen tensions along the abscissae and the percentage saturation along the ordinates as shown in figure 32 2. The curve is a rectangular hyperbola. Huffer obtained such a curve for hemoglobin from calculations based upon the law of mass action (see below).

If the foregoing procedures are carried out with blood instead of with a hemoglobin solution a different type of curve is obtained as shown in figure 32 3. It is doubly inflected or S-shaped.

The dependence of the oxygen saturation of hemoglobin in an aqueous solution upon the partial pressure of the gas is in accordance with the law of mass action which states that "the velocity of chemical change is proportional to the product of the concentrations of the reacting substances". In this case the reacting substances are reduced hemoglobin and oxygen. The reaction is reversible and is represented thus



It is evident that in the foregoing procedures the tension of oxygen in the hemoglobin solution, or in the plasma in the case of whole blood, came into equilibrium with the oxygen pressure of the atmosphere introduced into the tonometer. Then, the concentration of the dissolved oxygen must be proportional to the partial pressure of the gas to which the solution was exposed (p 367). So, if C_0 = concentration of O_2 , C_R = concentration of reduced hemoglobin and C_H = concentration of oxyhemoglobin, then the velocity of the reaction of Hb with O_2 to form HbO_2 will be proportional to the product of C_0 and C_R multiplied by a constant k , and the reverse reaction, the dissociation of oxyhemoglobin (HbO_2) to reduced hemoglobin

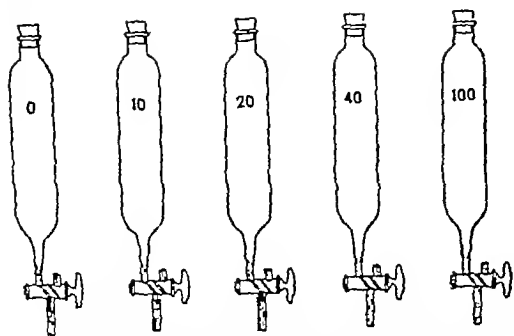


FIG 32.1 Series of tonometers. The numbers denote the pressure of oxygen in mm. (After Barcroft.)

(Hb) and oxygen, will be proportional to C_H multiplied by another constant k_1 . Thus

$$k(C_O \times C_R) = k_1(C_H)$$

At any given tension of oxygen in the opposed reactions, the formation of oxyhemoglobin and its dissociation proceed simultaneously until equilibrium becomes established. A simple reaction of this nature will explain the dissociation curve of a hemoglobin solution but the S-shaped curve of blood is thought to be due to a series of reactions. It will be recalled that natural hemoglobin (p. 57) has a molecular weight (68,000) four times greater than was previously supposed. It may therefore be given the symbol Hb_4 . It is thought that each molecule of hemoglobin combines with four molecules of oxygen, $Hb_4 + 4O_2 \rightarrow Hb_4O_8$. The oxygen dissociation curve calculated by the application

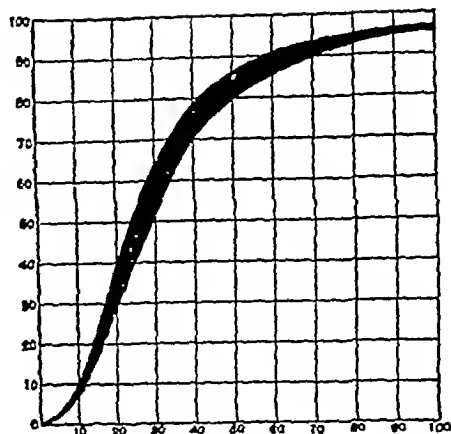
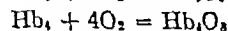
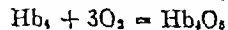
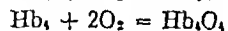
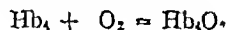


FIG 32.3 Composite curve to show the degree of variation in the oxygen dissociation curve of human blood as determined upon a number of persons. In each case the blood was exposed to an atmosphere containing the same percentage of CO as that of the alveolar air of the individual examined. Ordinates, percentage saturation, abscissae, oxygen pressure. (After Barcroft.)

of the law of mass action to this equation has a more pronounced S shape than any actually observed for blood. On the other hand, the curve calculated from the equation $Hb_4 + O_2 \rightleftharpoons Hb_4O_2$ is hyperbolic (fig. 32.2). It is believed, therefore, that in the combination of oxygen with hemoglobin under physiological conditions, four separate but simultaneous reactions take place:



The combination of all these reactions, it has been suggested, is responsible for the special S-shape of the oxygen dissociation curve (figs. 32.3 and 32.4).

Certain features associated with the shape of the dissociation curve of whole blood as shown in figure 32.3 are of the utmost physiological importance. It will be seen that with the partial pressure of oxygen in arterial blood (97 mm Hg) the hemoglobin is already nearly (98 per cent) saturated with the gas. Exposing the hemoglobin to a higher oxygen tension, therefore, will cause but a small increase in the total quantity of oxygen taken up by the blood, and to gain even the ultimate 2 or 3 per cent the oxygen pressure would require to be raised to over 300 mm Hg. The flattening out of the upper part of the curve means that relatively little reduction in the percentage saturation of the

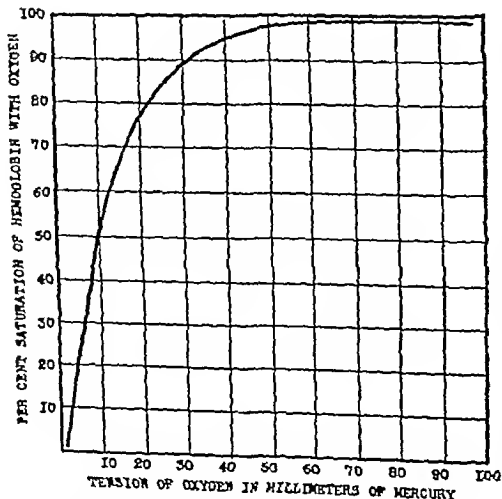


FIG 32.2 Oxygen dissociation curve of a solution of pure hemoglobin. (After Barcroft.)

hemoglobin occurs until the oxygen pressure falls to about half its normal value. At pressures above 60 mm Hg, relatively wide variations in alveolar oxygen pressure can occur with minimal changes in the total oxygen load of the blood. At an oxygen tension of 70 mm Hg the hemoglobin is still about 90 per cent saturated. The slope of the lower part of the curve is such that a given fall in oxygen pressure causes a much greater desaturation of the hemoglobin. The behavior of hemoglobin as indicated by the shape of the curve therefore favors a nearly maximum uptake of oxygen in the lungs so long as the oxygen pressure is above 80 mm Hg and a rapid release of the gas at the lower oxygen pressures which prevail in the tissues (1 mm or less to 60 mm Hg). It will be realized from a glance at the hyperbolic curve shown in figure 29.9 how unsuitable hemoglobin would be as a carrier of oxygen if it behaved in the manner indicated by such a curve. The hemoglobin would show a great avidity for oxygen in the lungs but would not yield up its oxygen load until the partial pressure in the tissues had fallen to a very low level. At the pressures which exist in the tissues the rate of dissociation of oxyhemoglobin would be many times slower than the rate of its formation. Hemoglobin would thus be worthless as an oxygen carrier. As it is, the oxygenation of hemoglobin in the lungs and its reduction in the tissues proceed at practically equal rates.

THE EFFECTS OF REACTION AND OF TEMPERATURE UPON THE SHAPE OF THE OXYGEN DISSOCIATION CURVE

The effect of dissolved inorganic salts upon the shape of the dissociation curve appears to be of less importance than was previously supposed. Hemoglobin solutions free from salts give the S-shaped curve when special precautions are taken to prevent a change (denaturation) occurring in the natural constitution of the hemoglobin complex. Denaturation may result from bacterial action, "aging" or other agencies.

A change in the reaction of the blood toward the acid side causes the curve to flatten toward the right, i.e., the affinity of hemoglobin for oxygen is reduced. Carbon dioxide and lactic acid liberated during tissue activity will exert this effect. The influence exerted by CO_2 was discovered by Bohr and is usually referred to as the Bohr effect¹.

¹ The oxygen dissociation curve of hemoglobin is so sensitive to acid that it has been used by Barcroft and

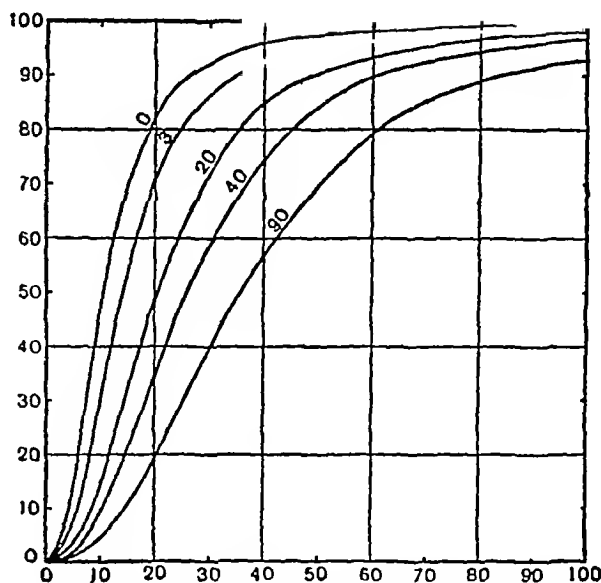


FIG. 32.4. Dissociation curves of human blood, exposed to 0, 3, 20, 40 and 90 mm CO_2 . Ordinates, percentage saturation. Abscissae, oxygen pressure (After Barcroft.)

(see figure 32.4) *Temperature* exerts a similar effect upon the dissociation curve. These agencies, therefore, cause the hemoglobin to liberate its oxygen more readily at the lower oxygen tensions but exert little effect upon the uptake of oxygen at the higher tensions. They cause the reactions involved in the breakdown of oxyhemoglobin to be speeded up, the equilibria shown on page 372 being shifted to the left. Alkalis and a fall in temperature will of course have the reverse effect.

THE UNLOADING OF OXYGEN IN THE TISSUES

It is now possible to give a summary of the manner in which oxygen is taken up from the lungs by the blood and supplied to the tissues. The absorption of oxygen from the alveolar air is the result simply of diffusion (pp. 365 and 369). The oxygen tension in the arterial blood is lower (never higher) than that of the alveolar air. This is so, even at high altitudes where the alveolar oxygen tension is greatly reduced (p. 424). There is no reason, therefore, to believe that the pulmonary epithelium actually secretes oxygen into the blood.² The per-

his associates as a basis of a method for measuring changes in blood reaction. A shift of the curve to the right, i.e., when the blood absorbs less oxygen at a given O_2 pressure, would indicate an increase in H^+ ion concentration, and a shift to the left the reverse change.

² The view that the pulmonary epithelium does not always play the rôle of a passive membrane but is capable of actively secreting oxygen into the arterial blood was advanced by Bohr and later elaborated by Haldane.

centage saturation of the hemoglobin is dependent upon the oxygen tension of the blood, i.e., the amount of the gas in simple solution, which, as just mentioned, is governed in turn by the oxygen pressure in the alveolar air. We have seen that the hemoglobin in the blood leaving the lungs is about 98 per cent in the form of oxyhemoglobin (i.e., it is 98 per cent saturated with oxygen).

Practically no oxygen is lost from the arterial blood until the capillaries have been reached, so the blood reaches the periphery with a high pressure head of oxygen. The oxygen tension of the tissue fluids and cells is relatively low, but varies with the functional activity at the moment between 1 and 60 mm Hg, a flow of oxygen from the plasma across the capillary membrane results. This, of course, will tend to lower the oxygen tension of the capillary blood plasma and upset the equilibrium between it and the oxyhemoglobin. Dissociation of the latter occurs to sustain the partial pressure of the plasma oxygen. In other words a steady flow of oxygen from red cell to tissue cell is maintained as a result of the slope in partial pressure of the gas. The rise in temperature and liberation of carbon dioxide and lactic acid in the tissues, as we have seen, shifts the oxygen dissociation curve to the right and so accelerates the decomposition of the oxyhemoglobin.

COEFFICIENT OF OXYGEN UTILIZATION

Depending upon the particular tissue and its activity at the moment, the blood in its passage through the capillaries loses from a fifth to nearly all of its oxygen load. The quantity of oxygen given up to most tissues during rest is about one quarter of that with which it leaves the lungs. That is, the venous blood coming from a tissue such as muscle during rest has an oxygen saturation of around 75 per cent, the equivalent of an oxygen partial pressure of about 40 mm Hg. During very heavy work the hemoglobin may give up nearly all its oxygen store to the contracting muscles. The mixed blood coming to the lungs, i.e., blood from all organs of the body, during rest is somewhat more than 75 per cent saturated with oxygen, containing about 45 volumes per cent less than does the arterial blood. The figure used to express the proportion of the total oxygen content of the blood which is given up to the tissue is called the *coefficient of oxygen utilization*. Thus, if the oxygen content of the arterial blood is 19.5 volumes per cent and that of the venous blood 15 volumes

per cent (i.e., an arterio-venous oxygen difference of 4.5 volumes per cent) the coefficient is $\frac{4.5}{19.5} =$

0.22. As indicated above, the coefficient varies considerably for different tissues and for the same tissue in accordance with the degree of its activity.

THE RESPIRATORY QUOTIENT

The ratio of the volume of carbon dioxide produced by a tissue to the volume of oxygen absorbed

$$\frac{\text{Volumes CO produced}}{\text{Volumes O}_2 \text{ absorbed}}$$

is called the respiratory quotient of that particular tissue. The ratio of these volumes as determined from the expired air will be the respiratory quotient of the body as a whole.

THE MANNER IN WHICH THE CALL OF THE TISSUES FOR OXYGEN IS MET

Increased activity of any tissue always entails an increased oxygen consumption which may be several times the resting value. Increase in the oxygen supply to the tissue above its requirement on the other hand does not increase the oxygen usage. The tissue takes what oxygen its activity at the moment demands but no more.

There are two ways in which a greater demand of the tissues for oxygen may be met. By (1) increasing the total blood flow through the tissue and maintaining a high intracapillary oxygen pressure, and (2) raising the coefficient of oxygen utilization, i.e., increasing the quantity of oxygen abstracted from a given volume of blood.

Both these factors come into play but the extent to which each operates is not the same for different tissues. The coefficient of oxygen utilization is increased by establishing a steep oxygen pressure gradient between the plasma within the capillaries and the tissue cells. That is, the quantity of oxygen used (Q) will, other things being equal, be proportional to the difference in intracapillary and intracellular pressures (P_C and P_T respectively), thus

$$Q \propto P_C - P_T$$

The pressure gradient of oxygen from blood to tissue cells acts in a sense as a force to drive oxygen through the tissues. The farther the cells are from the source of supply—the blood—the lower the

oxygen pressure will be and the less oxygen will they receive. During activity or any pronounced fall in oxygen tension of the blood, the cells most distant from the source of supply may suffer anoxia unless more capillaries open up, and thus reduce the radius along which oxygen diffuses (fig 32 5). Some tissues, however, such as cardiac muscle, take up oxygen at a pressure as low as 5 mm Hg (Keilin). An increase in the pressure gradient is brought about, (a) by the action of acids (carbonic and lactic) and a rise in temperature both of which accelerate the decomposition of oxyhemoglobin and so maintain a high intracapillary oxygen pressure, (b) by lowering the oxygen tension within the tissue cells, and (c) by shortening the distance through which the oxygen must diffuse, i.e., by opening up more capillaries and so reducing the radius of the cylinder of tissue supplied by each capillary (fig 32 5).

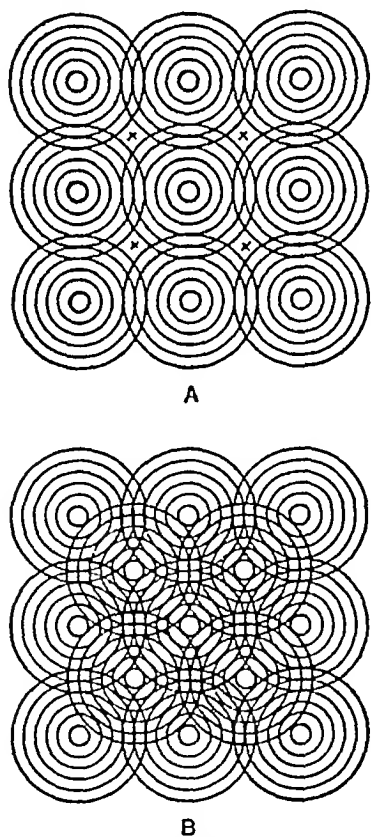


FIG 32 5 The pressure of oxygen in the tissues. The dot in the center of each series of concentric circles represents a capillary containing blood with an oxygen pressure of 30 mm Hg. Between each circle and the next larger one there is a fall of 5 mm Hg in oxygen pressure. At the points marked X (in A) the oxygen pressure is therefore zero, and the tissue is asphyxiated. In B, extra capillaries have opened up and there is a finite oxygen pressure throughout the tissue (After Barcroft.)

During activity more oxygen is consumed and the oxygen tension in the tissues tends to fall, thus increasing the pressure gradient from blood to tissue cells.

As mentioned above, during arduous work the coefficient of oxygen utilization is much increased, and, since the blood flow through the muscles is greatly increased, as well, the oxygen supply to the contracting muscles is augmented several fold. If, for example, a two-fold increase in the coefficient of oxygen utilization occurred together with a four-fold increase in blood flow the actual amount of oxygen delivered to the contracting muscles would be increased ($2 \times 4 =$) 8 times. In man the oxygen consumption of the muscles may actually be increased by from twenty to thirty times or in an athlete even fifty times. It has been found that the resting gastrocnemius of the cat consumes on the average 0.003 cc of oxygen per gram of muscle per minute. During activity the blood flow increases some 6 times, and the oxygen consumption over seven-fold. The increase in oxygen consumption must be brought about in part by the removal of more oxygen from each unit of blood.

The circulation rate (ch 26) in man may increase ten-fold or more during muscular exercise. It is certain, however, that there is also a redistribution of the blood, that is, a greater *proportion* of the total blood volume is driven through the muscles than during rest. The actual blood flow through the active muscles may, for this reason, be twenty times or more greater than the flow during rest. Nevertheless, though the red cell has a shorter stay in the capillary it unloads more of its oxygen, as indicated by the increase in the coefficient of oxygen utilization. The actual speed of the red cell through the capillary is not, however, as great as might be thought from the increased volume of blood flowing through the muscles, for the opening up of more capillary channels and the consequent increase in the total cross section of the blood stream (pp 140 and 178) tends to prevent any excessive acceleration of the red cell's passage.

BIOLOGICAL OXIDATION

By A. M. WYNNE

INTRODUCTION

Living cells and the complex organisms built up by the association of cells display characteristics of orderliness and integration, and of stability of structure combined with constant metabolic activity, of a type unmatched by any inanimate structure. The structure of

the cell, with its multiplicity of surfaces, makes possible a constant and orderly interplay and progression of a great variety of enzymatically catalyzed chemical reactions of a highly special kind which take place within the boundaries of the cell and are coordinated with reactions in neighboring cells and fluids. At ordinary temperatures and in media that are nearly neutral, living organisms can catalyze the transformation of chemical compounds in such a manner that their potential energy is made available to the cells for the performance of work of different kinds including the synthesis of specific cellular constituents, of widely varying complexity, from unspecific and relatively simple materials. These materials are taken from the environment and are transformed into a complex structure having a pattern and composition so characteristic of the species to which the organism belongs that not only are they regularly repeated in single individuals but they also emerge in an essentially unchanged form in many successive generations. This apparent constancy and stability of structure is not, however, to be regarded as a manifestation of a static condition, for we know that many of the actual material substances in the complex structure are in a state of constant metabolic flux, undergoing continuous degradation and resynthesis.

While this stability of structure is very great it is, however, not absolute, as evolution shows, and though the structural pattern and chemical composition are essentially the same in individuals of the same species there is evidence that in their fundamental chemical composition these individuals may differ slightly but significantly from one another (*cf.* Leo Loeb, 1945).

It is through the ordered and integrated activity of the many enzymatic catalysts of the cell that the existence of the living state is maintained, and the highly organized heterogeneous structure of the cell is undoubtedly concerned pre-eminently with the perfect coordination of the myriads of chemical reactions which constitute the total metabolism of the organism. After the death of the organism enzymatic activity continues for a while and chemical reactions still take place, but they are predominantly disintegrative in nature. The ability to synthesize specific cellular substances including proteins is lost, and the integrated activities of the living organism give way to processes of degradation.

The main purpose of this section of the chapter is to inquire into the nature and mode of action of some of the enzymes concerned with intracellular oxidation. These enzymes are components of the catalytic systems that exercise control over the oxidative reactions which provide the organism with a large proportion of the energy which it uses. Another important result of the action of these enzymes is, of course, the formation of numerous intermediary compounds which may serve as starting points for a great variety of metabolic syntheses. In the first part of the discussion an attempt will be made to describe in rather broad outline the principal types of oxidizing enzymes and their mode of

action in relation to the oxidation-reduction systems of the cell which are responsible for the transport of hydrogen and of electrons from the metabolites to molecular oxygen. Later in the chapter reference will be made to studies of oxidative condensation reactions, the over-all effect of which may be regarded as a "coupling" of an oxidative process with a condensation. The best known example of such a coupled reaction is oxidative phosphorylation, which results in the formation of adenosine triphosphate (ATP). This compound plays an important role in the metabolic activity of the living organism and is known to take part in reactions leading to the synthesis of a great variety of cellular substances.

In trying to understand the complex intermediary changes which matter undergoes in passing through living organisms the biochemist has found it expedient, as a part of his work, to adopt the practice of "pulling the organism to pieces" and of studying the catalytic activity of tissue slices and of cell fragments (in the form of enzyme preparations of varying degrees of 'purity') in promoting chemical transformations *in vitro*. The value and the defects of this general method of investigation are easily recognized on the one hand it has advanced our knowledge of the catalytic potentialities of the organism as a whole and has enabled us to learn something about the nature of many specific catalytic systems concerned with individual phases of intermediary metabolism, but on the other hand it is open to the general criticism that results of studies carried out *in vitro* may not necessarily serve as a true reflection of the integrated metabolism of the living organism itself in which surface-effects, diffusion phenomena, ionic balance and other factors exercise such a dominant influence. Since much of our knowledge of the catalytic systems controlling the transformation of energy in biological materials has been gained in studies of the activity of isolated 'fragments' of the cell, it can be argued that attempts to translate the results of these studies into generalized interpretations of the dynamic behavior of the organism as a whole may be premature. To a very considerable extent this criticism is justified, for there are still wide gaps in our knowledge of the cell interior and, especially, of the relationship between intracellular metabolic reactions and the physicochemical properties of the cell. It should be emphasized, therefore, that although considerable progress has been made in outlining the general pattern of the catalytic systems that control processes of oxidation in biological materials much still remains to be learned concerning the operation of these systems in the integrated metabolism of the organism as a whole.

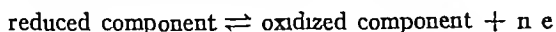
GENERAL PRINCIPLES OF BIOLOGICAL OXIDATION

Oxidations in general can be accomplished by the addition of oxygen, by the removal of hydrogen or by the withdrawal of electrons. All three methods are

fundamentally similar in principle, for, in each case, there is a loss of one or more electrons from the substance undergoing oxidation. Reduction, on the other hand, means, primarily, a gain of electrons, though the actual mechanism of reduction may involve the addition of hydrogen or the loss of oxygen. In biological systems the direct union of molecular oxygen with cellular metabolites has never been satisfactorily demonstrated although oxygen is, of course, the ultimate oxidizing agent in all aerobic organisms. While it is true that oxygen may be added to the molecule, as in the oxidation of an aldehyde to an acid, it is not molecular oxygen that is added, but, rather, oxygen that has its origin in some other source such as water or phosphate. In the case of the majority of the metabolites concerning whose physiological oxidation anything definite is known, the primary oxidation is effected by the removal of two hydrogen atoms (or two hydrogen ions and two electrons) through the agency of a specific dehydrogenase. It is with this hydrogen that oxygen eventually reacts, to form either water or hydrogen peroxide. The union of the liberated hydrogen with oxygen may be a direct one in the case of a relatively few dehydrogenases, such as certain flavoproteins whose activity leads to the formation of H_2O_2 . In other cases the liberated hydrogen atoms, or hydrogen ions and electrons derived from them, are believed to be transported to molecular oxygen by a series of reversible oxidation-reduction systems. The hydrogen is oxidized to water, and free energy is liberated. The molecular oxygen which takes part in oxidative reactions finally appears, therefore, only in the form of water, as the result of oxidation of hydrogen. No molecular oxygen enters directly into the formation of CO_2 , the oxygen contained in this product of respiration was either present originally in the compound oxidized or was added to it in the form of water or of phosphate, at some intermediary stage of oxidation.

OXIDATION-REDUCTION POTENTIAL

An oxidation-reduction (O-R) system contains two chemically related components which are capable of reversible transformation by the transfer of one or more electrons (or of H -ions and electrons). The relationship between the reduced and oxidized components (frequently called the 'reductant' and 'oxidant' respectively) may be represented by the equation



where n represents the number of electrons (e) transferred. The oxidizing or reducing power of such a reversible system in relation to that of other similar systems depends on two main factors: (1) a capacity factor which is governed by the concentration or, more accurately, the 'activity' of the reactants, and (2) an intensity factor which is a fundamental property

dependent upon the nature of the substances concerned. The intensity factor is the characteristic oxidation-reduction potential of the system and is represented by the symbol E_o in the standard equation

$$E_h = E_o + \frac{RT}{nF} \ln \frac{[\text{Oxidant}]}{[\text{Reductant}]}$$

where E_h is the potential (in volts) of an inert metallic electrode measured in a system of arbitrarily chosen ratio of [Oxidant] to [Reductant], and referred to the normal hydrogen electrode, at absolute temperature T . R is the gas constant, expressed in electrical units, n is the number of electrons transferred, and F is the faraday. E_o is the standard potential of the system at $pH = 0$, referred to the normal hydrogen electrode, when $[Ox] = [Red]$. When the concentrations of oxidant and reductant are equal the ratio is unity and the logarithm zero, then $E_h = E_o$. The normal hydrogen electrode ($pH = 0$) has been arbitrarily assigned a potential of zero. On increasing the pH , at constant temperature, the potential of the H -electrode decreases by approximately 0.06 volt for each unit of pH , at $pH 7.0$ its potential is -0.421 v. Similarly, the potential of other systems may be affected by changes in pH if, for example, the ionization of the reactants is influenced by pH -changes. And since the determination of the O-R potentials of biological systems is usually made at pH levels other than $pH = 0$, it is necessary to modify the term E_o in the above equation in conformity with this practice and with the effect of pH -changes. The observed O-R potential of a given system depends, therefore, not only on the innate tendency of the system to yield or accept electrons and on the ratio of the concentration of the oxidant to that of the reductant, but also, quite frequently, on the pH . In the equation

$$E_h = E'_o + \frac{RT}{nF} \ln \frac{[\text{Oxidant}]}{[\text{Reductant}]}$$

the term E'_o represents the characteristic potential of the system at the designated pH and its value is equal to that of the observed potential (E_h) at this pH when $[Oxidant] = [Reductant]$. Values of E'_o , measured at pH levels close to neutrality, have been determined for a considerable number of reversible biological systems, they vary from very negative values (e.g. -0.371 v for the system, hypoxanthine \rightleftharpoons xanthine at $pH 7.0$) to quite positive values (e.g. $+0.254$ v for the system, ferrous cytochrome C \rightleftharpoons ferric cytochrome C at $pH 1.7-7.7$). Lists of such values are recorded by Fischer (1939), by Oppenheimer and Stern (1939) and by Barron (1943b). Systems with more negative potentials are more reducing than those with more positive potentials. The H -electrode with a potential of -0.421 v at $pH 7.0$ represents a system of great reducing intensity, at the opposite extreme is the oxygen-electrode having a potential of $+0.810$ v at

pH 7.0, representing a system of great oxidizing intensity

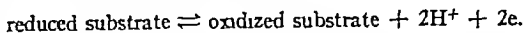
It should be emphasized that in order that an oxidizing or a reducing solution may give a definite, stable potential two important conditions must be satisfied. In the first place, the oxidized and reduced forms must *both* be present in definite amounts, the potential of a pure oxidant or of a pure reductant is meaningless. Secondly, each of the two components must be capable of conversion into the other as the result of an infinitesimally small change in the equilibrium or static potential in one direction or the other. This means that the system must be thermodynamically reversible, only the potentials measured in such systems have any fundamental significance and only such systems play any significant role in the physiological transport of hydrogen and electrons. Reference to the chemical nature of some of the physiologically important transport systems will be made later. Meanwhile it may be stated that of the several systems concerned some are most easily recognized as carriers of H-atoms or of H^+ ions and electrons, whereas others such as cytochrome transport electrons only. In all cases, however, the catalytic role of the carrier systems in the respiratory process depends on the exchange of electrons.

The significance of O-R potentials in studies of the mechanisms of biological oxidation is twofold. In the first place, knowledge of the characteristic potentials of the several systems which take part in the transfer of electrons from the metabolites to molecular oxygen permits an evaluation of the possibility of reaction between two given systems and provides information as to the sequence in which the several systems may co-operate in electron transfer. Under suitable conditions it is possible that a system with a more negative potential will reduce a system having a more positive or higher potential. The sequence of the interaction of the several O-R systems which constitute the respiratory mechanisms of living organisms is probably determined by many factors, one of which—obviously an important one—is the relationship existing between the characteristic potentials of the different reversible systems. But the fact that two O-R systems react with each other *in vitro* does not necessarily mean that they will also react *in vivo*. Considerations of this kind may impose serious limitations on the interpretation of the results of *in vitro* experiments in terms of intracellular reactions. Knowledge of the potentials of reacting systems has also been used in calculating the amount of free energy liberated at successive stages of the respiratory process. Suppose that a system of higher potential is reduced *in vitro* by one of lower potential, and that the difference between their characteristic potentials (E_0'), referred to the normal hydrogen electrode, is E (expressed in volts) when each system is half reduced (or oxidized). The free energy change ($-\Delta G$) is calculated from the

equation $-\Delta G = nEF$, where n is the number of electrons transferred and F is the faraday (i.e. approx. 23060 calories)

DEHYDROGENASES

A dehydrogenase catalyzes the oxidation of a specific metabolite (AH_2) by causing it to give up hydrogen (or H^+ ions and electrons) to a suitable H acceptor (B) which, thereby, becomes reduced $AH_2 + B \rightarrow A + BH_2$. Many dehydrogenases have been shown to catalyze reactions which are thermodynamically reversible and which, as *isolated* reactions, can be represented by the following equilibrium



The enzyme not only activates the reduced form (AH_2) of the substrate as a H-donor, but it also activates the oxidized form (A) as a H acceptor. A dehydrogenase-system represented by the equilibrium above is an oxidation-reduction system characterized by a definite O-R potential (E_0'), measurable *in vitro*, which depends upon the chemical nature of the two components and upon other factors such as pH. But the equilibrium is stable only if the ratio of the two components remains unchanged. It is unlikely that such an equilibrium can exist in the living cell because of the presence of other O-R systems, of lower or higher potential, which tend continuously to disturb the equilibrium by donating or accepting electrons.

Dehydrogenation of the reduced component (AH_2) of a reversible dehydrogenase system can take place only if there is also present another O-R system of higher, more positive, potential, so that its oxidized component (B) can accept hydrogen or electrons from the reduced form of the first system. The reaction of the two systems with each other is reversible and, in the absence of disturbing factors, will reach a dynamic equilibrium $AH_2 + B \rightleftharpoons A + BH_2$, with the result that the oxidation of AH_2 may be far from complete. The process of dehydrogenation of AH_2 can go to completion only if BH_2 can become a H-donor for another acceptor which is the oxidized component of an O-R system whose potential is higher than that of system B. In this way, through the mediation (as H^+ ion- and electron-carriers) of a series of O-R systems of gradually increasing potential (i.e. having increasing oxidizing intensities), the equilibrium, above, is constantly shifted to the right and the dehydrogenation of AH_2 progresses. Eventually the system represented by molecular oxygen, having a very high potential, comes into play and the transfer of hydrogen ions and electrons from the metabolite to O_2 is completed by a virtually irreversible series of reactions in which all of the free energy of the hydrogen is released.

In such a series of reactions this energy is liberated in a step-wise manner and energy is made available to the cell at each step in an amount which depends on the

difference of potential between the two systems which react with each other at that step. The total energy released in the several steps is equivalent to the energy liberated by the direct oxidation of the hydrogen to water by molecular oxygen.

Classification of dehydrogenases On the basis of present knowledge of their activities most of the known dehydrogenases can be classified tentatively according to the method of disposal of the hydrogen which is liberated from the substrate as the result of the catalytic action. Several such methods are now recognized.

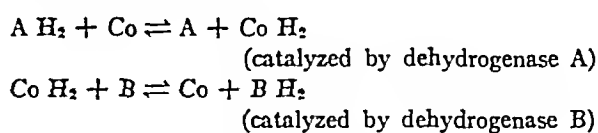
(1) The hydrogen is transferred directly to molecular oxygen, to form H_2O , by a prosthetic group attached to the enzyme protein, as in the case of certain flavoproteins such as D-amino acid oxidase. The peroxide formed in the reaction is subsequently removed by catalase or, possibly, by peroxidase. Water is the final product.

(2) The hydrogen is oxidized to water through the direct mediation of the cytochrome system between the dehydrogenase system on the one hand and molecular oxygen on the other. It has been customary to include among the dehydrogenases reacting in this manner a few enzymes such as succinic dehydrogenase, lactic dehydrogenase of yeast, α -1(+)-glycerophosphate dehydrogenase of muscle, and two or three others. But in the case of succinic dehydrogenase it now appears that an additional mediator is necessary (See Keilin and Hartree, 1949, and Slater, 1953).

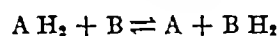
(3) The hydrogen is transferred to molecular oxygen, to form water, through the mediation of a more complex series of oxidation-reduction systems, including one of the two co-dehydrogenases (coenzymes, pyridine nucleotides), a flavoprotein, cytochrome, and cytochrome oxidase, functioning in the order named. Dehydrogenases reacting in this manner are more numerous than all other types and are further differentiated, according to the nature of the necessary pyridine nucleotide, into three groups: (a) those which require diphosphopyridine nucleotide, (b) those which require triphosphopyridine nucleotide, (c) one or two dehydrogenases which exhibit no specificity with respect to the pyridine nucleotide. Members of group (a) are considerably more numerous than those of the other groups.

Coenzyme-linked dehydrogenases In aerobic (respiratory) systems the hydrogen liberated from certain metabolites can be regarded as being transported in one direction to molecular oxygen by one or other of the above methods or by some similar method. The results are: (1) the metabolite is oxidized by the removal of hydrogen, and (2) the hydrogen is oxidized by union with oxygen. In glycolytic systems where anaerobic conditions prevail or where the supply of oxygen is insufficient to enable the respiratory processes to meet the demands of the cell for energy, the hydrogen released from the metabolite is not oxidized by molecular oxygen but may,

instead, be taken up by another, different, metabolite which is activated as a hydrogen-acceptor by its own specific dehydrogenase and thus becomes reduced. This means that two different dehydrogenase systems, each containing an enzyme and its specific substrate, react with each other in such a manner that the reduced component (A H_2 , below) of one system is oxidized by the oxidized component (B) of the other, while the latter is, of course, reduced to B H_2 . Several linked reactions of this kind have been demonstrated *in vitro* in the presence of one of the pyridine nucleotides (Co) which functions in the transfer of hydrogen, as illustrated in the following equations:

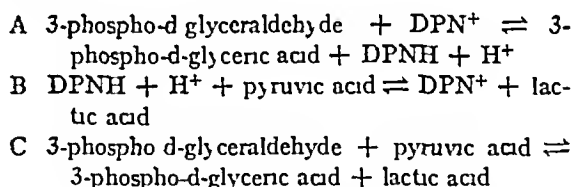


The net result is expressed as follows:

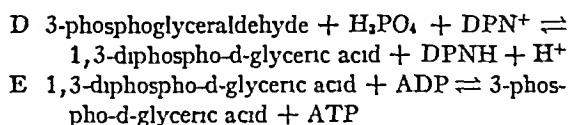


In glycolysis, and in fermentations in general, the oxidation of one compound is always accompanied by the reduction of another. Molecular oxygen does not, of course, take part in the reactions.

An important example of a coenzyme-linked reaction is the reduction of pyruvic acid by phosphoglyceraldehyde to form lactic acid, the end product of glycolysis in muscle.



Reaction C is a summation of reactions A and B but it is enabled to proceed only through the mediation of the coenzyme which functions as an oxidation-reduction system in the transfer of hydrogen. It should be stated that reaction A, as written above, is a simplification of the following reactions, D and E, in which 1,3-diphospho-d-glyceric acid is first formed and is later dephosphorylated enzymatically in the presence of adenosine diphosphate (ADP), to yield 3-phospho-d-glyceric acid and adenosine triphosphate (ATP).



Reaction D would seem to be a good example of an oxidation coupled with phosphorylation, for the aldehyde is oxidized (by DPN) to an acid, and at the same time phosphorylation of the carboxyl group appears to result from the uptake of inorganic phosphate. It will be shown later that this oxidative phosphorylation re-

action is somewhat more complex than is represented above, but the reaction illustrated is sufficient for our present purposes (See Racker, 1955)

CYTOCHROME AND CYTOCHROME OXIDASE

Let us turn now from the dehydrogenases, which can be regarded as functioning at one end of the respiratory chain, to the O R systems at the other end which are concerned with making molecular, inactive, oxygen available to the cell for the oxidation of hydrogen liberated from cellular metabolites. These latter systems constitute the cytochrome system consisting of the enzyme cytochrome oxidase and three hemochromogen like pigments, cytochromes a, b, and c. Each of these four compounds is believed to be a reversibly oxidizable protein-complex containing an iron-porphyrin as its prosthetic group, oxidation and reduction being effected in each case, by the exchange of an electron of the iron atom ($\text{Fe}^{++} \rightleftharpoons \text{Fe}^{+++} + e$). The discovery of the tetrapyrrol grouping in the structures of these compounds has served further to emphasize its special biological importance, for it is present also in chlorophyll (in which magnesium replaces iron), in hemoglobin, in catalase, in peroxidase and in other compounds of physiological significance. The tetrapyrrol structure appears, therefore, to be necessary not only to enable green plants to utilize solar energy in photosynthetic reactions and for transporting molecular oxygen from the lungs to the tissues of higher animals, but also for its utilization in energy-liberating reactions within the tissues themselves. The reader is referred to two extensive reviews one by Theorell (1947) on hemoproteins and the other by Granick and Gilder (1947) on the tetrapyrroles.

Warburg was the first to realize the importance of the iron porphyrin structure in intracellular oxidation. Since a large proportion, sometimes 90 per cent or more, of the normal oxygen-consumption of cells of many types can be blocked by small amounts of cyanide, Warburg was led, several years ago, to suspect that the enzyme chiefly responsible for the catalysis of oxygen-consumption in aerobic cells is a compound containing a heavy metal such as iron. Later, when he observed that the oxygen-consumption of the cells was greatly diminished in the presence of carbon monoxide if the system was kept in the dark, but that respiration was not affected by CO when the cell suspension was exposed to light, he suggested that the enzyme is a conjugated protein whose prosthetic group is an iron porphyrin. Still later, by means of an ingenious photochemical method, Warburg was able to obtain, as he believed, an accurate picture of the absorption spectrum of the CO-compound of the enzyme, without even attempting to remove the enzyme from the cells. The spectrum so obtained resembled very closely the spectra of the CO-compounds of several heme-containing substances such as pyridine and nicotine hemochromogens, these latter spectra were

determined not only by the conventional methods but also by the same photochemical procedure as had been employed with cell suspensions (See Warburg, 1949). It is not possible, in this short section, to include a description of Warburg's experimental procedures and observations, nor is it necessary to consider the theoretical aspects of his interpretations. He was, however, led to the definite conclusion that the enzyme concerned with cyanide sensitive respiration in living cells is an iron porphyrin protein complex, in spite of subsequent criticism of his methods and interpretations, his conclusion as to the chemical nature of the enzyme has never been seriously challenged.

To this enzyme Warburg gave the name "the oxygen-transferring enzyme", and, in the early formulations of his ideas as to its mode of action, he attributed to it the function of "oxygen-activation", as a result of which oxygen is enabled to oxidize directly the metabolites of the cell. This conception differed fundamentally from that of Wieland which emphasized the dominant rôle of the dehydrogenases as the activating agents responsible for the consumption of oxygen. Wieland regarded the liberated, "activated" hydrogen as undergoing direct oxidation by molecular oxygen.

As a result of experiments initiated some thirty years ago by Keilin, it became clear that neither the dehydrogenase systems, alone, nor an agent of the type of Warburg's enzyme, alone, can catalyze the oxidation of cellular metabolites by molecular oxygen. Keilin showed that both types of catalyst are essential components of a more complex oxidative system than that postulated either by Warburg or by Wieland. In 1886 MacMunn had observed in muscle and other tissues certain pigments which he named myo- and histo-hematin, until 1925, however, they had failed to arouse any widespread interest. In that year Keilin announced the discovery of three heme-containing pigments, in aerobic cells and tissues of many kinds, to which he gave the names cytochrome a, b, and c, the c-component appeared to correspond, spectroscopically, to the modified pigment of MacMunn. These pigments were later shown by Keilin to function as mediators linking the dehydrogenase systems of the cell with the system composed of molecular oxygen (O) and an enzyme, formerly known as indophenol oxidase but now generally recognized as cytochrome oxidase. This enzyme resembles very closely Warburg's "oxygen transferring enzyme" and is probably identical with it, but, instead of oxidizing directly a large number of cellular compounds in the manner suggested by Warburg, it is restricted in its action to the oxidation of the ferrous iron of one of the cytochromes in the presence of molecular oxygen.

The three pigments together with cytochrome oxidase constitute a system which transports electrons to O , and, by doing so, makes possible the oxidation of the hydrogen liberated from cellular compounds which

are activated by specific dehydrogenases. Though all four components of the cytochrome system are believed to be iron-porphyrin protein compounds, the oxidase differs from the cytochromes in several respects. For example, though the oxidase reacts directly with oxygen in the presence of cytochrome in the normal physiological pH-range, cytochrome c itself is autoxidizable only if the pH is below 4 or above 13. Cytochrome b can react sluggishly with oxygen under physiological conditions and to some extent, therefore, may be independent of the oxidase. Cytochrome a does not seem to be autoxidizable. At physiological pH-levels cytochrome oxidase in its ferrous form can combine with carbon monoxide to form a light-dissociable compound, and in its ferric form with cyanide, as a result of either combination the activity of the enzyme is inhibited, apparently by interference with the transfer of electrons. The reaction of cytochrome itself with CO is much less significant. Indeed, Keilin and Hartree (1939) could find no evidence of combination of cytochrome c with either CO or cyanide. Theorell (1947) has stated, however, that if the pH of the medium is sufficiently low (pH 2) or high (pH 13) the absorption bands of ferrous cytochrome c at 550 $m\mu$ and 520 $m\mu$ are shifted in the presence of CO to new positions at 563 $m\mu$ and 530 $m\mu$, respectively. At pH 1 or pH 14 the appearance of the new bands is much more rapid, but between pH 3 and pH 12 there is no evidence of the formation of a compound of CO with cytochrome C.

Though ferrous cytochrome oxidase in the tissues could be expected to combine with any CO that might reach the tissues it is unlikely that this reaction is ordinarily responsible for the toxic action of this gas. Instead, the toxic effect is mainly due to the fact that CO forms a difficultly dissociable compound with hemoglobin in the blood with the result that its O_2 -carrying capacity is greatly diminished. In the case of cyanide poisoning, however, it is the ferric form of cytochrome oxidase in the tissues that combines with cyanide, and the oxidase is thereby immobilized as an electron-transporter. Hemoglobin, normally present in the blood almost entirely in the ferrous form, does not combine with cyanide. It is only the ferric form, methemoglobin, that reacts with cyanide.

Though previously thought to be incapable of union with cyanide, ferric cytochrome c has recently been shown to form with cyanide *in vitro* a complex whose absorption spectrum differs from that of cytochrome c itself (see Horecker and Kornberg, 1946). Since complex-formation can take place at pH 7.4 one might suspect that this reaction would contribute to the toxic effect of cyanide on respiration *in vivo*, if it were not for the fact that the reaction of cyanide with cytochrome c *in vitro* is so very much slower than with cytochrome oxidase. The rapid immobilization of the oxidase by cyanide would have the effect, *in vivo*, of maintaining the cytochrome in the ferrous form after reduction by

the dehydrogenase systems of the cells, in this form cytochrome cannot combine with cyanide.

Several other cytochromes have been described in the biochemical literature as constituents of living organisms of different kinds, their identification have been based mainly on their spectral properties. These cytochromes with one exception, namely cytochrome a_1 , do not, however, concern us here. This compound, called a_2 by Keilin to distinguish it from a_1 and a_2 which are said to occur in certain bacteria, possesses spectral and other properties closely similar to those of Warburg's "oxygen-transferring enzyme" (cytochrome oxidase), and for this reason cytochrome a_2 has been assumed to be identical with cytochrome oxidase. Not only are the spectral characteristics of the two substances very similar, but both compounds are thermolabile and autoxidizable, and both combine with the commonly used respiratory inhibitors including carbon monoxide. But difficulty has been experienced in demonstrating the oxidation of reduced cytochrome c by oxidized cytochrome a_2 , and it has not yet been possible to cause dissociation of the CO cytochrome a_2 complex by exposure to strong light. For these reasons there may still be some doubt as to the complete identity of the two compounds cytochrome oxidase and cytochrome a_2 .

In his early studies of cytochrome Keilin demonstrated that substances such as narcotics which were known to inhibit the activity of cellular dehydrogenases inhibit also the reduction of oxidized cytochrome as revealed by spectroscopic observation. Other inhibitors of cellular oxidation, such as cyanide, which in very low concentrations specifically inhibit cytochrome oxidase (but in these concentrations have little or no effect on dehydrogenase activity), were found to inhibit also the oxidation of reduced cytochrome. Keilin concluded therefore that the primary function of cytochrome in cellular oxidations is the transport of electrons from the dehydrogenase systems of the cell to cytochrome oxidase which, in turn, reacts with molecular oxygen. It became evident from these and similar investigations that the cytochrome system as a whole constitutes a terminal mechanism which in many organisms reacts directly with molecular oxygen and in this way makes available to these organisms a large proportion of the oxygen which they consume.

Upon their release from the metabolites, or more usually from a flavoprotein which has served as an intermediary hydrogen-carrier, the "metabolic" hydrogen-atoms are ionized and yield H^+ -ions. The electron lost by each H -atom is believed to be taken up in succession by the trivalent iron of the components of the cytochrome system, in other words by four iron-porphyrin compounds, reducing each in turn ($Fe^{+++} + e \rightleftharpoons Fe^{++}$). Each cytochrome on yielding an electron to the ferric iron of the adjacent member of the chain becomes reoxidized and is then in a position to accept another electron. Finally, the divalent iron of the

oxidase itself, the last member of the series, is oxidized to the trivalent state by O_2 . The mechanism of this oxidation is still only imperfectly understood, without attempting to discuss the theoretical aspects of the problem, it is sufficient to state that the interaction of O_2 and the ferrous form of cytochrome oxidase in an aqueous medium facilitates the transport of electrons by the cytochrome system as a whole and makes possible the oxidation of metabolic hydrogen to water. It is unlikely that all four compounds necessarily function in the cyanide-sensitive respiration of aerobic cells of all types, since some types of cells are known to lack one or other of the three cytochromes. The sequence in which the four compounds participate in the transfer of electrons to O_2 is believed to be as follows: cytochrome b \rightarrow cytochrome c \rightarrow cytochrome a \rightarrow cytochrome oxidase. This conclusion is based upon considerations of the OR potentials of the several systems, E'_0 (pH 7.0) being -0.04 v, $+0.254$ v and $+0.290$ v for cytochromes b, c, and a, respectively.

Only one of the pigments, namely cytochrome c, has been obtained in soluble form, the others are so intimately bound to the insoluble material of the cells that they have resisted all attempts to separate them from this material. The c-component has, however, been highly purified. The purest preparation obtained by Theorell had an iron content of 0.43 per cent and molecular weight of 13000. The structure and properties of this product have been described by Theorell and Åkesson (1941) and by Theorell (1947). It is of considerable interest that approximately 30 per cent of the total nitrogen (including the porphyrin nitrogen) of cytochrome c is represented by the essential amino acid lysine, a compound which has not yet been identified with any specific reaction or function in living organisms. No other protein is known to contain such a large proportion of lysine. The distribution and certain metabolic aspects of hemoproteins, including cytochrome c, have been discussed by Drabkin (1948).

Very little is known about the chemical nature of cytochrome oxidase other than the well established fact that it contains an iron-porphyrin as its prosthetic group. The details of the structure of the iron porphyrin (or "heme") are, however, not known. There is some evidence, based on considerations of the positions and intensities of the absorption bands identified by photochemical methods, that the heme present in the cytochrome oxidase of yeast differs spectroscopically from that in a similar enzyme in rat heart, suggesting possible structural differences, for example, in the side-chains attached to the porphine ring. The heme of the yeast enzyme seems to fall in the class of the "mixed color" hemes or "pbeobemes" which in their spectral properties are intermediate between the red and green hemes. Nothing is known regarding the nature of the protein which is attached to the iron porphyrin in cytochrome oxidase. Warburg has recently computed the molecular weight of the enzyme to be 75,000 per

mole of heme, a figure which is roughly six times the molecular weight of cytochrome c (13,000).

THE PYRIDINE NUCLEOTIDES AND FLAVOPROTEINS

Reference has been made to the fact that many dehydrogenases depend for their activity upon the co-operation of a coenzyme which is either diphosphopyridine nucleotide (Co I, DPN) or triphosphopyridine nucleotide (Co II, TPN). The coenzyme is regarded by some workers as a prosthetic group bound to a specific activating protein or "apodehydrogenase", the two components together forming a "pyridinoprotein" having the properties of a specific dehydrogenase. Others prefer to look upon the activating protein itself as the dehydrogenase and the coenzyme as one of its substrates. At the moment of catalysis the metabolite and the coenzyme are both bound to the activating protein, and hydrogen is transferred from the metabolite to the coenzyme, with the result that the former is oxidized and the latter reduced. Unless, however, the reduced coenzyme can be reoxidized, the extent of oxidation of the metabolite is negligible since it is limited by the small amount of coenzyme present.

The coenzymes themselves do not react directly with molecular oxygen. Until a few years ago the only known physiological agents which could be shown to effect their oxidation at adequate speeds *in vitro* were two specific flavoproteins. One of these, called cytochrome c-reductase and found in yeast, catalyzes specifically the oxidation of reduced coenzyme II (TPNH) by cytochrome c and in this way may link the cytochrome system with the dehydrogenase systems of yeast which depend upon TPN. The other flavoprotein, named diaphorase or coenzyme-factor, has been extracted from animal tissues, yeast and bacteria and has been shown to catalyze the oxidation of DPNH (and possibly TPNH) by methylene blue *in vitro*. In an aerobic system the resultant leucomethylene blue is reoxidized by O_2 , so that a relatively small amount of the dye enables the physiological carriers to transport relatively large amounts of metabolic hydrogen to O_2 . The direct reaction of diaphorase with the cytochrome system has not been satisfactorily demonstrated, however, the facts that diaphorase can oxidize DPNH and that the oxidation, in tissue slices, of the substrates of several dehydrogenases which require DPN is sensitive to cyanide, suggest that diaphorase or some closely related flavoprotein intervenes between DPNH and the cytochrome system. Straub in 1939 obtained from pig heart a soluble flavoprotein having all the catalytic properties of the coenzyme-factor, the activity of which had hitherto been associated with insoluble particles in the tissue extracts. The activity of Straub's flavoprotein was such that under optimum conditions each molecule of the enzyme catalyzed the aerobic oxidation of ca. 8000 molecules of DPNH per minute by carriers such as methylene blue.

Two other enzymes in animal tissues have been shown recently to catalyze the reduction of cytochrome

c by DPNH and TPNH respectively. The first of these was found by Hogeboom (1949) in the mitochondria and submicroscopic particles of rat liver homogenates, and has been named DPN-cytochrome c reductase since it enables cytochrome c to oxidize reduced coenzyme I. The second enzyme, TPN cytochrome c reductase, prepared by Horecker (1949) in the form of a pale yellow solution by treatment of liver extracts and shown to be present in the mitochondrial fraction, is similar in its action to cytochrome c reductase of yeast since it reacts specifically with TPNH but not with DPNH.

Only a brief reference to the chemical nature of the coenzymes and flavoproteins is possible here. Both pyridine nucleotides (coenzymes) are, in reality, dinucleotides containing two mononucleotides united through their phosphoric acid residues. TPN contains a third phosphoric acid residue but is, otherwise, similar in structure to DPN. One of the mononucleotides in each coenzyme is adenylic acid (adenosine-5'-phosphate), the other mononucleotide contains nicotinic acid amide, d-ribose and phosphoric acid united in this order (i.e. nicotinamide d-ribose-5'-phosphate). The flavoproteins are compounds in which a specific activating protein is bound to a prosthetic group which may be either a mononucleotide, flavin phosphate, (i.e. the 5'-phosphate of 6,7-dimethyl, 9-d-ribityl isoalloxazine) or a dinucleotide, flavin adenine dinucleotide, containing flavin phosphate and adenylic acid united through their phosphoric acid residues.

Reactions A and B outlined on page 384 illustrate the manner in which hydrogen, released from cellular metabolites through the agency of specific dehydrogenases, may be transferred to a flavoprotein by way of diphosphopyridine nucleotide. But these two transporting systems can function catalytically, only if the reduced flavoprotein can be reoxidized so that it can then oxidize the reduced coenzyme once again. The oxidation of the flavoprotein may be effected in a manner similar to that suggested in scheme C on page 384, in which DPN-cytochrome c-reductase (an iron-containing flavoprotein) is pictured as transferring electrons singly through one of its iron atoms to the cytochrome system and thence to molecular oxygen. Hydrogen ions formed from the metabolic hydrogen ($2H \rightarrow 2H^+ + 2e$) are withdrawn from the environment, and the final result is the oxidation of the hydrogen to water.

Several flavoproteins have been shown to contain one or more metals (Fe, Cu, Mo) as integral constituents (see Mahler and Green, 1954). DPN-cytochrome c reductase is said to contain 4 atoms of iron in each enzyme molecule of molecular weight 80000. This iron is considered by Mahler and Elowe (1954) not only to facilitate the free flow of electrons from the reduced flavin moiety of the flavoprotein to the trivalent iron of cytochrome, but also to serve as a structural link between the flavin-adenine dinucleotide and protein components of the reductase. The failure of diaphorase

to react directly with cytochrome may be related to its comparative lack of iron. If DPN-cytochrome c reductase is rendered iron-free by treatment with metal binding agents, its ability to react with cytochrome is lost but its diaphorase activity (reaction with dyes) is retained.

The preceding brief description of some of the oxidative catalysts of living organisms does not, of course, embrace all oxidases which are believed to exist in living cells. But the oxidation systems which have been discussed appear to constitute a recurring pattern in organisms of widely different types. These respiratory enzyme systems are made up of two main components: (1) a specific catalytic protein whose function is concerned with the release of H-atoms from the oxidizable substrate, (2) a number of reversible oxidation-reduction systems which transport either the hydrogen atoms themselves or electrons which are derived from them, from the substrate to molecular oxygen. The potentials of these systems become more positive as the oxygen end of the respiratory chain is approached. In some cases only a single O-R system intervenes between the substrate and O_2 , as in the oxidation of L-amino acids by a specific flavoprotein. In the more complicated respiratory systems at least three different types of O-R systems take part in the transport of hydrogen and electrons, namely, one of the pyridine nucleotides, a flavoprotein, and the cytochrome system (respiratory). At each step along the respiratory path leading from the metabolite to O_2 free energy may be made available to the cell, the amount of energy depending partly upon the relative oxidation-reduction potentials of the adjacent O-R systems which react with each other in the transport of electrons. Of the total free energy liberated during the passage of one electron from the pyridine nucleotide to O_2 by way of flavoprotein and the cytochrome system, a large fraction, it would seem, is liberated during the passage of the electron through the cytochrome system to O_2 .

CATALASE AND PEROXIDASE

As the result of the activity of certain flavoproteins (e.g. D-amino acid oxidase) that transfer hydrogen direct from their substrates to molecular oxygen, hydrogen peroxide is formed. This product, a toxic substance, can be decomposed to water by catalase or by peroxidase. Until recently it was generally believed that the function of catalase is only a protective one, being concerned specifically with the degradation of H_2O_2 to water and O_2 , whereas peroxidase can act as a secondary oxidizing catalyst to promote the oxidation of cellular metabolites by H_2O_2 . It appears, however, from the work of Keilin and Hartree (1945) that catalase can also catalyze the secondary oxidation of certain compounds, such as alcohols, by H_2O_2 formed in the

idases do not. Both enzymes contain an iron-porphyrin prosthetic group which resembles very closely the corresponding group in hemoglobin. Any readers who may be interested in recent investigations of catalase and peroxidase are referred to Theorell's reviews (1947 and 1951)

GLUTATHIONE

This tripeptide, glutamyl cysteinyl glycine, discovered in 1921 by Hopkins, is almost universally distributed in the intracellular fluids of animal tissues as well as in plants, yeasts and many bacteria. Of all the known cellular compounds containing the sulphhydryl group, glutathione is probably the most widely distributed in living cells, for this reason and because it is capable of reversible oxidation to the disulphide form, many attempts have been made to demonstrate its activity as a hydrogen-transporting system in cellular oxidations. It is doubtful, however, whether such a function can be attributed to the compound, for all attempts to demonstrate the full reversibility of the change from the reduced (sulphydryl) form to the oxidized (disulphide) form have failed. The potential of the system appears to be determined solely by the reduced form, it is unlikely that a substance having these properties can play any important role in the physiological transport of hydrogen or electrons in respiratory systems. Moreover, the fact that it is present in living cells almost entirely in the reduced form suggests either that the reduction of its oxidized form is much more rapid than the oxidation of its reduced form, or that there is no cycle of reduction and oxidation.

Other functions of glutathione have, however, been more firmly established, for example its rôle in maintaining the —SH groups of enzymes. The presence of the sulphhydryl group in several oxidizing enzymes, as well as in several proteolytic and other enzymes, has been shown to be necessary for their activity (Bersin 1935, Hellerman 1937, Barron and Singer 1945, Elliott, 1946). In view of the fact that oxidizing agents in the cell tend to inhibit the activity of —SH enzymes, it is probable that one of the main functions of glutathione in cellular systems is concerned with the continuous reactivation of the —SH enzymes. Its capacity to bring about such a reactivation, through its powerful reducing action, has been demonstrated in many experiments.

It is not improbable that the enzyme "glutathione reductase" may have important physiological significance in this connection. This enzyme has been shown to be present in several different animal tissues and in other materials such as yeast and plant tissues, and it catalyzes the reduction of oxidized glutathione (GSSG) by TPN-dependent dehydrogenase systems such as the system consisting of glucose-6-phosphate dehydrogenase and its substrate. Moreover, it has been demonstrated (Rall and Lehninger, 1952) that partially purified preparations of the reductase are able to catalyze the oxidation of reduced TPN (but not of

reduced DPN) in the presence of GSSG, as followed spectroscopically by observing the change in optical density at 340 m μ . The enzyme is therefore believed to catalyze the reaction



As this reaction is practically irreversible, it would seem that glutathione reductase may be an important agent in maintaining glutathione in the reduced form under physiological conditions.

RELATION OF VITAMINS TO OXIDATIVE CATALYSTS

Four water-soluble vitamins, nicotinic acid, riboflavin, thiamine and pantothenic acid, give rise *in vivo* to physiologically active derivatives which are closely related to intracellular oxidations. Nicotinic acid amide is a constituent of the pyridine nucleotides, DPN and TPN, riboflavin enters into the formation of the flavo-proteins. These two vitamins are necessary therefore for the formation of compounds which are primarily concerned with the transport of hydrogen or of electrons along the respiratory chain from cellular metabolites to molecular oxygen. A third vitamin, thiamine, is converted by phosphorylation *in vivo* into thiamine pyrophosphate (TPP, diphosphothiamine, co-carboxylase), a compound which appears to be an essential component of the enzyme systems which catalyze the oxidative decarboxylation of α -keto acids such as pyruvic acid and α -ketoglutaric acid. Other essential components of these systems are Mg^{++} ions, inorganic phosphate DPN, coenzyme A and lipoic acid. Coenzyme A is a derivative of the water soluble vitamin, pantothenic acid, and is believed to possess the structure shown in figure 32 6A, in which are incorporated one residue, each, of β -mercapto ethanolamine, pantothenic acid, D-ribose and adenine, and three residues of phosphoric acid. Frequently used forms of abbreviation are HS-CoA and CoA, the first of these emphasizes the important functional —SH group the presence of which permits the formation of compounds such as the acyl S-CoA derivatives (e.g. acetyl-S-CoA).

Though not usually included among the vitamins, lipoic acid (thioctic acid) is another biocatalyst whose participation in the reactions resulting in the oxidative decarboxylation of α -keto acids appears now to be well established. Knowledge of the biochemical activity of this compound has been gained mainly in studies of oxidations and transfer reactions catalyzed by bacteria and by animal tissues. During the progress of investigations of its isolation and chemical nature and of its presence and activity in organisms of widely different types ranging from protozoa to mammals and from bacteria to higher plants it has been variously known as Factor II, acetate-replacing factor, protogen, and pyruvate oxidation factor (POF). Now, however, each of these "factors" appears to be identical with lipoic acid, the structure of which is shown in figure 32 6B. Closely related isomers such as the 4,8- and 5,8-dithio acids have been shown to be less active than lipoic

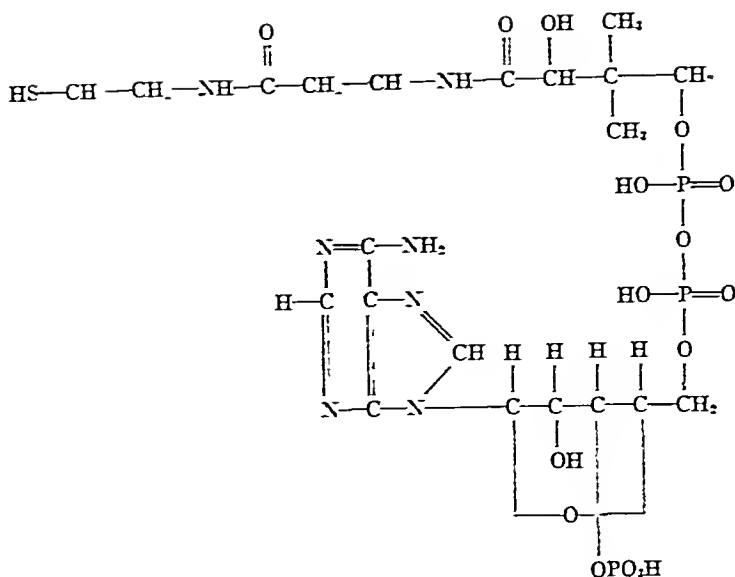


FIG 32 6A. Provisional structure of Coenzyme A (after Novelli 1953)

acid as coenzyme (POF) in the oxidation of pyruvate by cells of the lactic acid organism *Streptococcus faecalis* harvested from a lipoic acid-free medium. Reference will be made later to the manner in which lipoic acid participates as a coenzyme, together with TPP, DPN and HS-CoA, in the oxidative decarboxylation of α keto acids, with the formation of free CO₂, reduced DPN and an acyl-S-CoA.

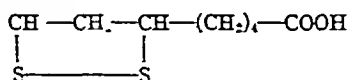


FIG 32 6B Lipoic acid (6,8-dithio-octanoic acid)

THE CITRIC ACID CYCLE

Meanwhile attention is drawn to figure 32 6C showing the pathway of the oxidation of carbohydrate and fat by way of the citric acid cycle. From both carbohydrate and fat, in preliminary reactions which are only partially shown, the same compound acetyl-S-CoA is formed, the two carbons of the acetyl group representing a fragment of the original carbohydrate or fat. A portion of this acetyl-S-CoA may enter into synthetic and exchange reactions, the remainder is oxidized to CO₂ and water. The first step in this oxidation is the reaction in which acetyl-S-CoA unites with oxaloacetate under the influence of the "condensing enzyme", to form citrate and at the same time to liberate the coenzyme in the free form (HS-CoA). The coenzyme may then undergo further reaction in one of the preliminary steps or at a later stage, in the conversion of α -ketoglutarate into succinyl-S-CoA. In each turn of the citric acid cycle the free oxaloacetate is regenerated and is thereby made ready for the oxidation of another molecule of acetyl-S-CoA. The operation of the cycle thus provides a catalytic mechanism for the oxidation of large amounts of the two-carbon fragment, formerly described as "active

acetate" but now known to be the acetyl group of acetyl-S-CoA derived from carbohydrate by way of pyruvate and from fatty acids by way of the related β -keto acyl-S-CoA compounds.

Adequate amounts of oxaloacetate are necessary to ensure the operation of the cycle with maximum efficiency. And because oxaloacetate tends to undergo spontaneous decarboxylation it is important that mechanisms for the synthesis of this substance be maintained. These mechanisms include the formation of oxaloacetate (1) from aspartate by transamination, (2) from other compounds such as malate, citrate and succinate, ingested with the food, from α ketoglutarate derived from glutamate by transamination and by oxidative deamination and from other amino acids by way of glutamate, (3) by condensation of CO₂ with pyruvate, and (4) by the joint operation of the "malic enzyme" (No 12) and malic dehydrogenase (No 11). Methods 3 and 4 depend upon available supplies of pyruvate, the main source of which is carbohydrate. If the production of normal amounts of pyruvate should fail, the supply of oxaloacetate will tend to diminish. As a result, the oxidation of acetyl-S-CoA may be retarded, and in the liver this substance may by conjugation give rise to large amounts of acetoacetyl-S-CoA which in turn can be hydrolysed in the liver to acetoacetate and free coenzyme. If, therefore, the efficiency of the operation of the cycle is seriously impaired, excessive amounts of "ketone bodies" may be formed, resulting in ketosis, ketonuria and acidosis.

In certain reactions of this cyclic mechanism and in some of the preliminary reactions which lead to the formation of acetyl-S-CoA from fat and carbohydrate, the pyridine nucleotide coenzymes (DPN and TPN) become reduced. Their ability to function catalytically as coenzymes depends upon their rapid reoxidation.

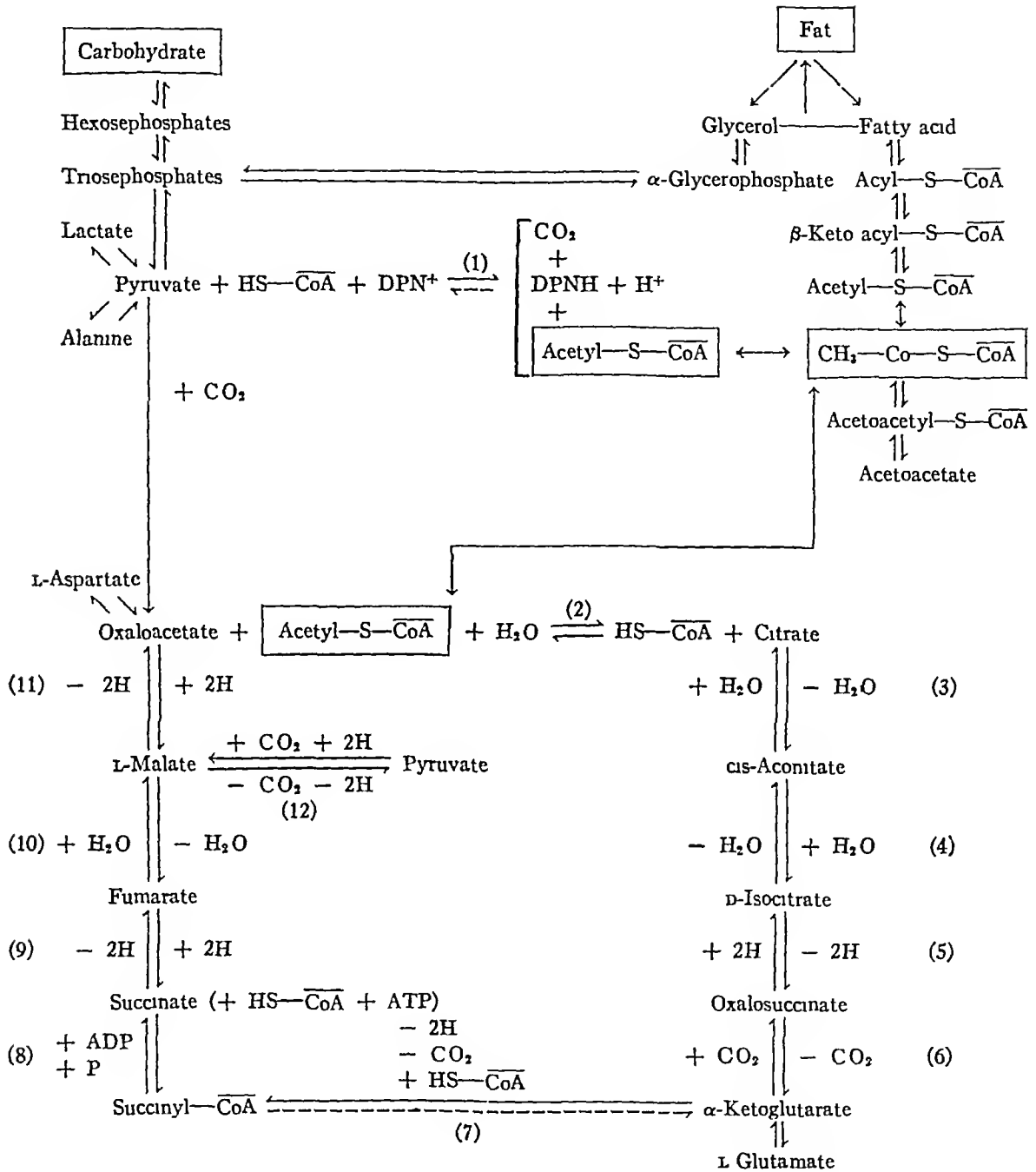


FIG 32 6C Outline of pathway of oxidation of carbohydrate and fat by way of the citric acid cycle

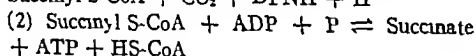
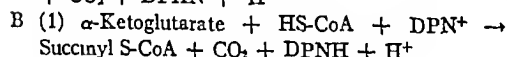
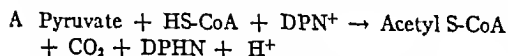
Enzymes

- 1 Pyruvic dehydrogenase ("oxidase") + HS-CoA, DPN, TPP and Lipoic acid
- 2 Condensing enzyme
- 3 Aconitase
- 4 Aconitase
- 5 Isocitric dehydrogenase + TPN
- 6 Oxalosuccinic decarboxylase
- 7 α-Ketoglutaric dehydrogenase + HS-CoA, DPN, TPP, and probably Lipoic acid
- 8 Succinyl-CoA phosphokinase
- 9 Succinic dehydrogenase
- 10 Fumarase
- 11 Malic dehydrogenase
- 12 "Malic enzyme"

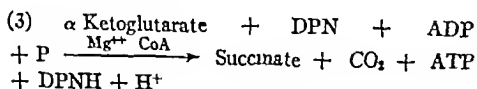
This reoxidation is effected in the manner previously outlined, through the joint action of the cytochrome c reductases and the cytochrome cytochrome oxidase system. An extensive discussion of the enzymic mechanisms of the citric acid cycle will be found in the review by Ochoa (1954).

The pyruvate and α ketoglutarate oxidation systems

The oxidative decarboxylation of pyruvate and α ketoglutarate is catalyzed, in each case, by a complex enzyme system containing, as previously noted, at least four essential coenzymes, namely, TPP, DPN, CoA and lipoic acid, together with inorganic phosphate (in the second case at least) and Mg^{++} ions as other essential components. Knowledge of these catalytic systems has accumulated over a period of years and is based on studies of the degradation of α keto acids by animal tissues and bacteria and by cell free enzyme preparations. The details of these studies can not be reviewed here, but if it were possible to do so it would be realized that only partial descriptions of the over-all reactions can be given. In each case DPN plays its customary role of hydrogen acceptor, an acyl-S-CoA derivative is formed (acetyl S-CoA in the case of pyruvate and succinyl S-CoA in the case of α ketoglutarate), and CO_2 is liberated. The reactions may be illustrated as follows:



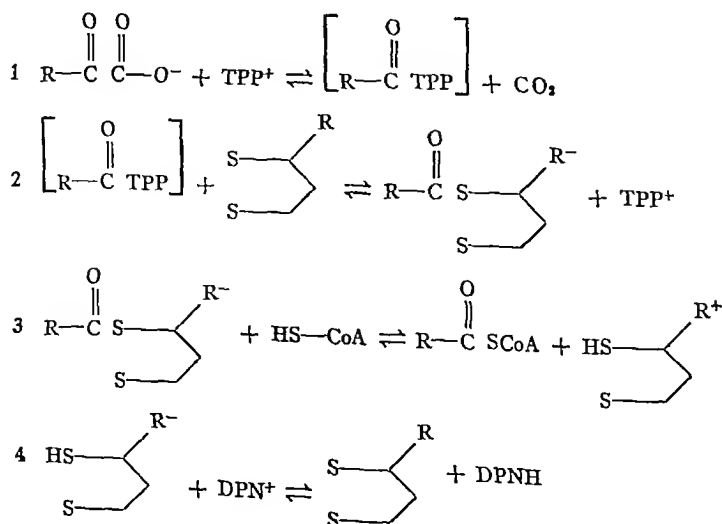
Sum of (1) and (2)



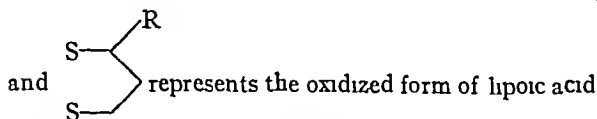
The acetyl S-CoA formed in reaction A undergoes metabolism in the manner previously discussed, and the succinyl-S-CoA formed in reaction B (1) may react as in B (2) with ADP and inorganic phosphate (P) under the influence of a specific enzyme, in the presence of Mg^{++} ions, to yield free succinate, ATP and free coenzyme A. It will be observed that in reaction B (1) a condensation (the synthesis of succinyl S-CoA) is coupled with the oxidation and in reaction B (2) the synthesis of ATP (another condensation) is coupled with the liberation of free succinate and coenzyme A from succinyl S-CoA. The net result shown in B (3) is the phosphorylation of ADP to ATP, coupled with the oxidative decarboxylation of α -ketoglutarate. In addition to the mechanism indicated in reaction B (2) there is evidence that succinyl S-CoA can be hydrolyzed enzymatically to yield free succinate and free coenzyme A.

Reactions A and B as formulated above make no reference to thiamine pyrophosphate or to lipoic acid, both of which are believed to be necessary in many species for the oxidative decarboxylation of α keto acids. It is interesting to note that of the several coenzymes which are essential components of the complete enzyme system the one concerning whose mode of action least is known is thiamine pyrophosphate, despite the fact that it was recognized many years ago as a coenzyme in pyruvate metabolism, long before coenzyme A and lipoic acid attracted the attention of investigators in this field.

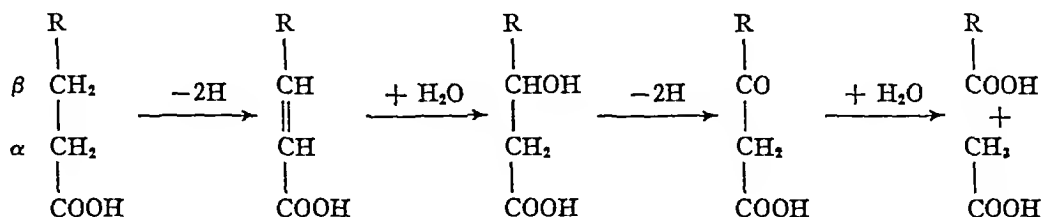
Gunsalus (1953, 1954) has suggested a mechanism which would bring into play all four coenzymes in a series of four reactions resulting in the oxidative decarboxylation of an α keto acid and the formation of an acyl-condensation product. The suggested reactions are shown below:



R represents, for example, $\text{CH}_3\text{—HOOC—}(\text{CH}_2)_2\text{—etc}$,



Of the four reactions included in this scheme, reactions 3 and 4 have been clarified, and the enzyme catalyzing reaction 4 has been separated from the others. The exact mechanism of the first two reactions is, however,



not yet clear. In the first reaction CO_2 is removed and thiamine pyrophosphate is pictured as forming a "carbanion-TPP" (aldehyde-TPP) complex. In the second reaction lipoic acid acts as a carbanion acceptor forming a thioester, the lipoic acid being thereby reduced. In the third reaction this thioester acts as an acyl donor to form acyl-S-CoA and reduced (or dimer-capto) lipoic acid. And in reaction 4 the reduced lipoic acid is re-oxidized by DPN. The four reactions pictured in this scheme should accomplish in succession the following results: 1) decarboxylation, 2) acyl generation, 3) acyl transfer, and 4) hydrogen transfer. Lipoic acid, either in free or in combined form, is concerned with the last three of these processes, while thiamine pyrophosphate is indicated as the initial aldehyde-acceptor and decarboxylating coenzyme in reaction 1.

Reed and DeBusk (1953) have presented evidence which they interpret as indicating that in a mutant strain of *Escherichia coli* the active coenzyme-form of lipoic acid is not the free acid but, rather, a conjugated form called "lipothiamide pyrophosphate" (LTPP) in which thiamine pyrophosphate through its primary amino group is joined by amide linkage with lipoic acid through its carboxyl group. This coenzyme is said to be associated, in sequence, with enzymic reactions resulting in (1) acyl generation accompanied by the liberation of CO_2 , (2) acyl transfer, with the formation of acyl-S-CoA, (3) hydrogen transfer. This series of reactions parallels almost exactly those in the scheme suggested by Gunsalus. But there is as yet no good evidence that lipoic acid and thiamine pyrophosphate exist generally in the conjugated form (LTPP) in living organisms.

OXIDATION AND SYNTHESIS OF FATTY ACIDS

The well-known theory of the β -oxidation of fatty acids, first proposed by Knoop in 1904 and supported by the independent work of Dakin shortly thereafter, has for a half-century engaged the attention of many biochemists. But only recently has it been possible to provide experimental confirmation of the essential

aspects of the theory by isolating and characterizing the several enzymes which appear to be responsible for the step-wise degradation.

It will be recalled that the classical scheme of fatty acid oxidation, outlined by Dakin and by Knoop and known as β -oxidation, included the following steps and provided for the formation of free acetic acid and for the oxidation of the β -carbon of the original acid to a carboxyl group.

The resulting acetate was presumed to undergo oxidation to CO_2 and water, and the R—COOH , representing the remaining portion of the original fatty acid, could then, in a similar series of reactions, give rise to another molecule of acetate and thereby be reduced in chain-length by two carbon atoms. Successive β -oxidations would, according to the theory, continue until a four-carbon residue of acetoacetate remained. This substance would normally be oxidized in the peripheral tissues and would not accumulate. And, according to the theory, acetoacetate should arise only from the four terminal carbons at the methyl end of the chain.

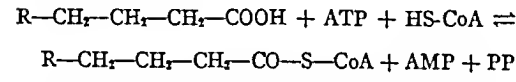
When later, as the result of the work of numerous investigators, it became evident that these terminal carbon atoms could not possibly represent the sole source of acetoacetate under certain conditions, modifications of the original theory were introduced suggesting that two-carbon fragments detached in the process of β -oxidation could condense with each other to form acetoacetate. The study of the metabolism of isotopically labelled fatty acids has provided convincing proof that condensation of this type can take place in animal tissues.

There remained, however, at least two serious problems which had to be solved before the general principle of β -oxidation could be accepted as representing a biochemical mechanism having important physiological significance. The first was the difficulty of reconciling with the theory the repeated failures of workers in this field of investigation in their attempts to isolate intermediary products of the oxidation of long chain fatty acids. The second problem was underlined by the great difficulties encountered in attempts to obtain cell-free enzyme preparations capable of catalyzing the oxidative processes. The difficulties inherent in the first problem became more readily understandable when the second problem was solved, with the isolation, purification and characterization of five individual enzymes which are now believed to be responsible jointly for the catalysis of a series of five reactions leading to the formation of

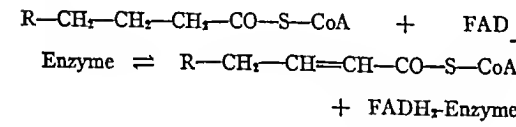
acyl S-CoA. And it has now become apparent that the intermediary acids can never exist in the free form because at every stage of the oxidative process coenzyme A is bound to the reactant, forming a transient intermediary whose concentration at any moment is severely limited by the catalytic amounts of available CoA

It is now believed that the fatty acid oxidizing capacity of liver cells is concentrated wholly in the mitochondria. It is unlikely however that any significant accumulation of an acyl CoA intermediary in the intact mitochondrion could occur because the product of the action of each of the five enzymes would be subjected immediately to the action of the next enzyme in the sequence. These five mitochondrial enzymes, shown in figure 32 6D, which catalyze the conversion of a fatty acid molecule into acetyl S-CoA have been brought into soluble form and have been highly purified. The reactions which they catalyze have been studied individually and are formulated as follows (Mahler 1953, Green 1954)

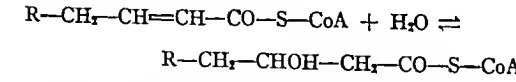
1 Formation of fatty acyl CoA by the fatty acid activating enzyme



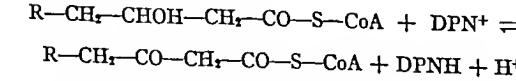
2 Formation of unsaturated fatty acyl CoA by the fatty acyl CoA dehydrogenase (metallo-flavoprotein)



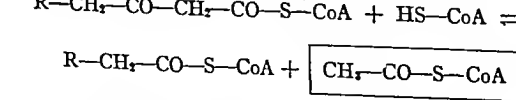
3 Formation of β -hydroxy acyl CoA by the $\alpha \beta$ unsaturated acyl CoA hydratase



4 Formation of β -Keto acyl CoA by the β hydroxy acyl CoA dehydrogenase



5 Formation of acetyl CoA by the cleavage enzyme



As a result of the activity of these enzymes functioning in the sequence shown, a "two carbon fragment" is removed from the fatty acid in the form of the acetyl group of acetyl CoA, and an acyl CoA is formed whose

chain length is shorter by two carbons than that of the original fatty acid. This newly formed acyl CoA can now give rise to another molecule of acetyl CoA and a third acyl CoA by way of reactions 2 to 5. Eventually the entire fatty acid is degraded to acetyl CoA. As we have already observed, the oxidation of the acetyl group of this compound is catalyzed by the enzymes of the citric acid cycle. The free HS-CoA is regenerated in the first (or condensation) reaction of the cycle and the liberated coenzyme A can now take part in reaction 1 again, in which another fatty acid molecule undergoes conjugation with the coenzyme to form the corresponding acyl CoA. This initial reaction requires ATP but the mechanism of the reaction has not yet been elucidated.

The complete degradation of a long chain fatty acid to acetyl CoA can be represented in the following manner (Green, 1954)

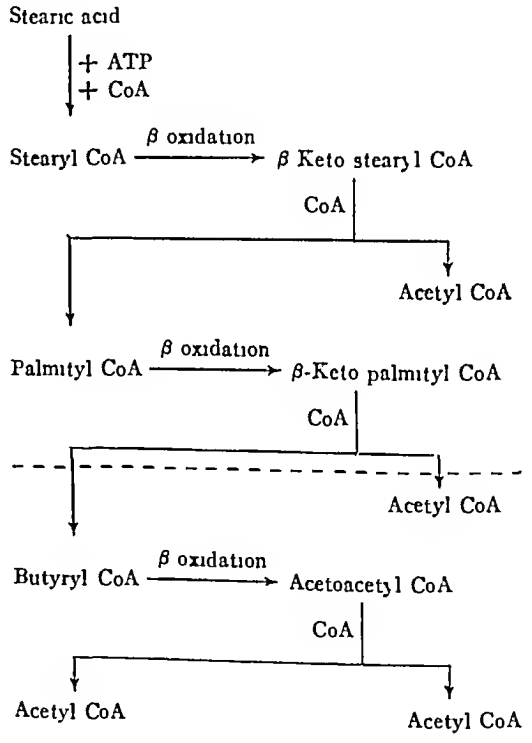


Figure 32 6D illustrates a probable pathway for the synthesis of fatty acids by way of the five reactions, E, D, C, B, A operating in the reverse (counter clockwise) direction. The completion of the series of reactions from E to B has the effect of elongating the carbon chain of a fatty acyl CoA derivative by two carbons. These carbons are contributed by the acetyl group of acetyl CoA which is condensed with the fatty acyl CoA by the reverse action of the cleavage enzyme (E). This initial condensation is followed in succession by the

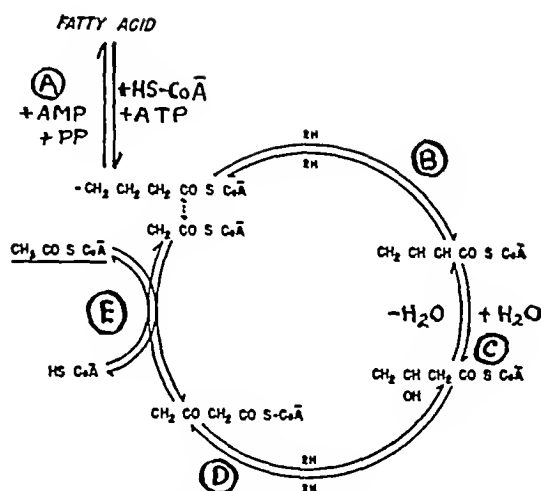


FIG 32 6D Oxidation and synthesis of fatty acids
(Adapted from Lynen, 1953)

Oxidation clockwise, synthesis counter-clockwise
Enzymes (terminology of Mahler, 1953)

- A Fatty acid activating enzyme
- B Acyl-S-CoA dehydrogenase (metallo-flavoprotein)
- C Alpha-beta-unsaturated acyl-CoA hydratase
- D Beta-hydroxy acyl-CoA dehydrogenase + DPN⁺
- E Cleavage enzyme

addition of two atoms of hydrogen, the removal of a molecule of water and the addition of another pair of hydrogen atoms. The new fatty acyl CoA, longer by 2 carbons than its precursor, can condense with another molecule of acetyl CoA in reaction E, and by a repetition of reactions D, C and B another fatty acyl CoA molecule is formed having two additional carbon atoms in its chain. In this manner the fatty acyl chain is increased in length by two carbons at a time and eventually reaches a maximum length which is determined by unknown factors. By the reversal of reaction A the free fatty acid is released from its combination with coenzyme A, and ATP is formed by the condensation of adenylic acid and inorganic pyrophosphate. The results of studies of the synthesis of long chain fatty acids from acetate labelled with C¹⁴ in the carboxyl group appear to be consistent with this theory of the mechanism of the synthetic process.

OXIDATIVE CONDENSATIONS AND THE PATHWAYS OF BIOLOGICAL SYNTHESIS

One of the major problems of biochemistry is the elucidation of the pathways and enzymatic mechanisms of the synthesis of the great variety of chemical substances which enter into the composition of living cells. These compounds can in many cases be regarded as products of the condensation of smaller units joined together by "anhydro" linkages in enzymatically catalyzed reactions involving the elimination of water. Polysaccharides, nucleotides, nucleic acids, peptides, glycosides, glycerides are familiar examples of conden-

sation products of this general type. It is well known that in the presence of water and a suitable catalyst these compounds can be broken down into smaller units in reactions which involve the addition of the "elements" of water (H and OH). Moreover, studies of the equilibrium constants of hydrolytic reactions of this kind catalyzed by specific enzymes have shown that in many cases the equilibrium between the condensed molecule and its split-products lies far over on the side of hydrolysis. Thermodynamic principles suggest that appreciable synthesis by direct reversal of hydrolysis would not be expected to occur except perhaps under conditions such as those permitting the continuous and effective removal of the condensed product from the system. Frequently, therefore, we find that for the synthesis of these condensation products enzymatic mechanisms are available which are quite different from the reversal of hydrolysis.

The free energy change of a reaction

In considering the problems of biological oxidations and syntheses it is desirable to refer briefly to the so-called free energy change (ΔG) of a reaction. All reactions which proceed spontaneously are accompanied by a decrease in the free energy of the system (ΔG is negative) and reactions of this kind are said to be exergonic. The *standard* free energy change (ΔG°) of a reversible reaction is usually expressed in calories and is related as follows to the equilibrium constant, K.

$$\Delta G^\circ = -RT \ln K \quad (\text{and at } 25^\circ\text{C}, \Delta G^\circ = -1365 \log K)$$

The relation between the *actual* free energy change (ΔG) of a reaction and ΔG° is

$$\Delta G = \Delta G^\circ + RT \ln$$

Product of activities of reactants on right hand side
Products of activities of reactants on left hand side

It is evident therefore that only when the products of the activities of the reactants on the two sides of the equation are equal is ΔG equivalent to ΔG° . If one wishes to calculate the free energy change of a reaction under conditions other than those of the standard state (which specifies molal concentrations of reactants and products) it is necessary not only to know the value of ΔG° but also to have accurate information regarding the concentration of reactants and products. And because information of this kind pertaining to the intact living organism is not easily obtained, calculations of the free energy change of reactions *in vivo* have frequently no precise value.

In some circumstances it may be possible to arrive at an approximate value, or range of values, for the ΔG of a reaction in the tissue itself, from a consideration of the concentrations of reactants and products found to be present by analysis of the excised tissue. For example, while the free energy of hydrolysis of ATP at pH 7.5 under thermodynamic standard conditions (1 e molal

concentrations of ATP, ADP and inorganic phosphate) is -9400 calories, Burton and Krehs (1953) have calculated ΔG of hydrolysis of ATP in animal tissues to vary from -13000 to -16000 calories, depending upon the concentrations of reactant and products. But they emphasize the fact that these values must be regarded as only rough estimates of ΔG since they are based on only a relatively few analyses. And inherent in the calculations is the assumption, perhaps unwarranted, that ATP, ADP and inorganic phosphate are uniformly distributed throughout the tissue.

It should be emphasized, in this connection, that reactions *in vivo* do not reach a state of thermodynamic equilibrium, but that the living organism is characterized by a "steady state" or dynamic equilibrium in which synthetic processes are balanced by processes of degradation. And while thermodynamic considerations may help us to understand the direction which reaction may follow *in vivo* and perhaps to assess the magnitude of the free energy changes in some cases, the free energy change of a given reaction and the efficiency of metabolic processes in a steady state system will frequently depend on the nature and on the rates of accompanying chemical reactions (see Hearon 1951, 1952). In studying the energetics and the economy of the living cell it is necessary therefore to strive to understand the kinetic factors which operate in the steady state system. Precise knowledge of these factors is at present very limited indeed.

"Phosphate bond energy"

Studies of the free energies of hydrolysis of phosphorylated compounds have led to an arbitrary division of these compounds into two general categories, characterized in the one case by relatively high values and in the other by relatively low values of ΔG° (hydrolysis). In the first group are included acyl phosphates such as 1,3-diphosphoglyceric acid, pyrophosphates such as ATP, enol phosphates such as phosphoenolpyruvate, and guanidino phosphates such as creatine phosphate. The values of ΔG° (hydrolysis) of these compounds vary from approximately -9000 to -15000 calories. In the second group are included phosphate esters such as α -glycerophosphate, glucose-1-phosphate, glucose-6-phosphate, fructose-1-phosphate, fructose-6-phosphate, 2-phosphoglyceric acid, 3-phosphoglyceric acid, values of ΔG° (hydrolysis) for these compounds fall in the range of approximately -2300 to -4800 calories (see Burton and Krebs 1953).

Compounds in the first group were described by Lipmann (1941) as containing "energy-rich" phosphate bonds, designated by the symbol $\sim\text{ph}$, whereas compounds of the second general type were said to contain "energy poor" phosphate bonds ($-\text{ph}$). Energy generated in catabolic processes was said to be "stored" in the energy rich phosphate bond and this "phosphate-bond energy", particularly in ATP, was regarded as a unique source of energy to be used in a great variety of cellular activities and syntheses.

Serious objection to this concept of "energy rich" phosphate bonds has been taken by Gillespie *et al* (1953) who emphasize the physico-chemical principle that the formation of a chemical bond when atoms combine to form molecules is accompanied by a *decrease* in the total energy of the system, and that the dissociation of a molecule into its constituent atoms (i.e. the disruption of chemical bonds) requires an expenditure of energy.

This is, of course, in direct contrast with the concept of energy *stored* in a chemical bond such as a phosphate bond, and for this reason a certain amount of confusion has been associated with the use of the term phosphate bond energy. Perhaps the danger of confusion is best avoided if the "energy richness" of phosphorylated compounds and, indeed, of all other classes of condensed molecules, is interpreted strictly in terms of ΔG° (hydrolysis) values. These values provide a measure of an important biochemical property of such compounds, namely their potential, in the presence of an appropriate enzyme and by virtue of their instability, to donate energy to systems at a lower energy level and at the same time to act as donors of one or other of their constituent moieties in reactions known as group transfer reactions. The expression "phosphate bond energy" as used by many biochemists today embraces these general principles.

Two distinct types of enzymic reaction are believed to be involved in the synthesis of a variety of naturally occurring condensed molecules (see Hanes 1953). These reactions are (1) oxidative condensation reactions leading to the synthesis *de novo* of primary condensed products, (2) group transfer reactions "in which one of the constituent groups of a condensed molecule, the donor, is transferred to new linkage with an acceptor."

The best understood of the oxidative condensation reactions is that catalyzed by triosephosphate dehydrogenase (see p. 379). The overall effect of this reaction was recognized as an oxidative phosphorylation, a "coupling" between the condensation of ADP and inorganic phosphate (P) to yield ATP, and the oxidation of 3-phosphoglyceraldehyde to 3-phosphoglyceric acid, accompanied by the reduction of DPN. The condensation of ADP and P is a highly endergonic reaction whereas the oxidation of 3-phosphoglyceraldehyde by DPN is strongly exergonic. In the complete "coupled" reaction the endergonic and the exergonic transformations become part of the same process which proceeds spontaneously, being accompanied by a relatively small decrease in the free energy.

It has transpired subsequently that the reaction is more complex than has been represented above. It now appears that the primary condensation product of the reaction is a 3-phosphoglyceryl mercaptide of a glutathione moiety which forms part of the active center of the enzyme, and that ATP is generated secondarily from this primary oxidative condensation product ($\text{R}-\text{CO}-\text{S}-\text{Enzyme}$) as a result of a sequence of two group-transfer reactions. The first of these involves

the transfer of the 3-phosphoglyceryl group from its attachment to the sulphur atom in the active center of the enzyme to inorganic phosphate, forming 1,3 diphosphoglycerate ($R-CO-O-PO_3H_2$) and liberating the HS-enzyme from its combination with substrate. Then follows a second reaction, catalyzed by a separate enzyme, in which the phosphate group is transferred from the carboxyl position of the 1,3-diphosphoglycerate to ADP to form ATP. Although the over-all process can be recognised as an oxidative phosphorylation, it would appear that phosphate is not involved in the actual oxidative condensation process.

Oxidative phosphorylation

It is now generally recognized that many oxidative reactions are accompanied by an uptake of inorganic phosphate, with the formation of organic phosphate or phosphates of the type of ATP. There appears therefore to be a coupling of these oxidative reactions with phosphorylation. But the exact nature of the products of the *primary* condensation reactions which accompany oxidative processes is, in most cases, unknown. There does not appear to be any good evidence that ATP is necessarily the primary product, ATP may, in most cases, be formed in secondary reactions involving group-transfer, as in the case of the oxidation of 3-phosphoglyceraldehyde. The over-all process can, however, be considered to be one of oxidative phosphorylation and there is much evidence to support the view that it represents a highly effective means of conserving the energy of oxidations.

The phosphorylations which are considered on the basis of available evidence to be possible during the complete oxidation of a triosephosphate molecule (representing one-half of a glucose unit) to CO_2 and H_2O are summarized below. The figures represent moles of inorganic phosphate taken up per mole of substance oxidized. $DPNH_2$ represents reduced DPN.

1 (a) 3-Phosphoglyceraldehyde + H_2PO_4 + DPN	
1,3 diphosphoglyceric acid + $DPNH_2$	1
(b) $DPNH_2 + \frac{1}{2}O_2 \rightarrow DPN + H_2O$	3
2 (a) 2-Phosphoglycerate \rightarrow phosphoenolpyruvate + H_2O (non oxidative)	0
(b) Phosphoenolpyruvate + ADP \rightarrow pyruvate \rightarrow ATP	1
3 (a) Pyruvate + oxaloacetate \rightarrow DPN \rightarrow citrate + CO_2 + $DPNH_2$	0
(b) $DPNH_2 + \frac{1}{2}O_2 \rightarrow DPN + H_2O$	3
4 (a) Isocitrate + TPN \rightarrow oxalosuccinate + $TPNH_2$	0
(b) $TPNH_2 + \frac{1}{2}O_2 \rightarrow TPN + H_2O$	3
5 (a) α -Ketoglutarate + HS-CoA + DPN \rightarrow CO_2 + $DPNH_2$ + succinyl-CoA	0
(b) Succinyl-CoA + ADP + P \rightarrow succinate + HS-CoA + ATP	1
(c) $DPNH_2 + \frac{1}{2}O_2 \rightarrow DPN + H_2O$	3

6 Succinate + $\frac{1}{2}O_2 \rightarrow$ fumarate + H_2O	2
7 (a) Malate + DPN \rightarrow oxaloacetate + $DPNH_2$	0
(b) $DPNH_2 + \frac{1}{2}O_2 \rightarrow DPN + H_2O$	3
Total per molecule of triosephosphate	20

In the oxidation of one molecule of glucose 40 phosphorylations would be effected in the reactions shown. Assuming that each phosphorylation forms a molecule of ATP and that 2 molecules of ATP are utilized in transforming a molecule of glucose into 2 molecules of triosephosphate, a total of 38 phosphorylations would accompany the oxidation of one glucose molecule, and if glycogen were the starting substance the number would be 39 per glucose unit in glycogen.

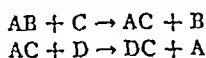
Reaction 2a is non-oxidative and there is no accompanying uptake of phosphate, but ATP is formed in 2b. Reactions 1a and 5b constitute phosphorylations "at the substrate level", while the phosphorylations which accompany the oxidation of $DPNH_2$ and $TPNH_2$ must occur during the passage of electrons through the flavoprotein and cytochrome systems to molecular oxygen. Though nothing is known concerning the chemical nature of the primary condensation products which are formed during the passage of electrons along the transport system, it is evident that the phosphorylations which accompany this passage of electrons represent a large fraction of the total number of phosphorylations associated with the oxidative process as a whole.

The figures shown above for reactions 1b, 3b, 4b, 5c, 6 and 7b represent also in these cases the ratio, moles of inorganic phosphate taken up/atoms of oxygen consumed in the process of oxidation. This is usually called the P/O ratio and it provides a measure of the extent of phosphorylation accompanying the different oxidative processes. The values of these ratios are based on results obtained in studies of reactions catalyzed *in vitro* by the appropriate enzymes, and are recorded as whole numbers after making allowance for some destruction of ATP by ATP-ase.

Of the seven groups of reactions shown above, the first two are part of the "glycolytic" series and are generally considered to be catalyzed by soluble enzymes in the cytoplasm. Reactions in groups 3 to 7, in addition to the coupled phosphorylations, probably take place in the mitochondria. And while the P/O ratios indicate the extent of phosphorylation which accompanies oxidations *in vitro*, it is difficult to interpret these results in relation to the highly integrated metabolism of the intact, undamaged cell. It is unlikely that there is an accumulation of ATP in amounts suggestive of a "pool" of this substance. It would seem more likely that the adenylic "system" consisting of catalytic amounts of ATP and ADP, together with inorganic phosphate, provides a means through which energy-production can, in a sense, be controlled by energy utilization.

Group-transfer reactions in biological syntheses¹

Reference has been made to the generation of ATP during the oxidation of 3-phosphoglyceraldehyde, as a result of two successive group transfer reactions. The essence of reactions of this type is that one moiety of a pre-existing condensed molecule (the donor) is transferred to new linkage with another molecule or group (the acceptor) without the intervention of hydrolysis, and in the process, the formation of a new anhydro bond is accompanied by the dissolution of a pre-existing one. A succession of two group-transfer reactions may lead to the liberation of both moieties of the original donor molecule and to the condensation of the two different molecules which acted as acceptors at the two stages. This is represented in the following manner:



The separate moieties A and B of the original donor AB are now available for resynthesis and the regenerated substance AB may then take part in a further cycle of reaction. In this way a condensed molecule (AB), synthesized in close association with oxidative condensation reactions, may serve as the "condensing agent" for the synthesis of cellular constituents.

Adenosine triphosphate occupies a special position as a condensing agent in synthetic processes, for it is now known to be an essential component of the specific enzyme systems which catalyze the synthesis of a great variety of substances. It is necessary, for example, for the phosphorylation of sugars and for the synthesis of glycogen from glucose, for the phosphorylation of compounds such as thiamine, pyridoxal, riboflavin, for the synthesis of peptides, of hippuric acid, and of glutamine, for the synthesis of acetyl choline, of citrulline from ornithine, and of arginine from citrulline, for transmethylation, as in the synthesis of creatine, and, in conjunction with coenzyme A, for the "activation" of fatty acids. Moreover, it plays a unique role as the medium through which energy is made available to contracting muscle (see Needham, 1952, for an extensive discussion of this role of ATP).

The part played by ATP in the synthesis of glycogen illustrates the principle of group-transfer. The addition of one glucose unit to the polysaccharide chain requires three separate transfer reactions, each catalyzed by a specific enzyme: (1) transfer of a phosphoryl group from ATP to glucose, to yield glucose 6-phosphate and ADP, (2) transformation of glucose 6-phosphate into glucose-1-phosphate in a group-transfer reaction involving glucose 1,6-diphosphate as an intermediary,

¹The writer is greatly indebted to Professor C. S. Hanes, F.R.S., for the opportunity of discussing with him the problems of biological synthesis. Readers are directed to reviews by Hanes, Connell and Dixon (1952) and by Hanes (1953) for further discussion of the significance of group transfer reactions.

(3) transfer of the glucosidyl group from glucose-1-phosphate to the polysaccharide chain, with the liberation of free phosphate. By repetition of this process glucose molecules are condensed to form polysaccharide. For every glucose unit added to the polysaccharide one molecule of ATP is broken down to ADP and free phosphate. These split products of ATP are available for resynthesis, and the condensation of glucose to polysaccharide may continue as long as oxidative condensation reactions are able to promote the resynthesis of ATP. Consideration of this and other pathways for the synthesis of complex cellular constituents shows that the condensing capacity of the cell would be indicated at any moment by the ratio [ATP]/[ADP] [P], and this in turn would reflect the relative rates of the breakdown of ATP (mainly, it may be presumed, by its participation as a donor in numerous group-transfer reactions) and of its resynthesis as a result of oxidative condensation processes. Reaction systems of this kind provide us with some understanding of the reasons underlying the dependence of various synthetic processes upon oxidation reactions.

REFERENCES—see p. 1357

THE CARRIAGE OF CARBON DIOXIDE BY THE BLOOD

By PROFESSOR J. K. W. FERGUSON

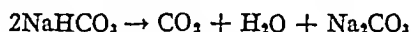
The total carbon dioxide content of blood means the amount of CO₂ which can be extracted from a given volume of blood by exposure to a vacuum after the addition of acid. The results are usually expressed as volumes per cent (v.p.c.) by which is meant cc. of CO₂ (measured at S.T.P.) per 100 cc. of blood. The methods of Van Slyke and his coworkers are now most commonly used for the measurement of CO₂ in blood and other fluids. In Van Slyke's manometric apparatus a known volume of blood is acidified and subjected to a partial vacuum in which CO₂ and other gases are rapidly extracted. The gases are then compressed to a known volume and their pressure measured. The CO₂ is removed by introducing NaOH and the pressure at the same volume is again measured. From the change in pressure the amount of CO₂ can be calculated. By such a method then, it is found that human blood normally contains 50 to 60 v.p.c. of CO₂, the venous blood usually containing 5 to 10 v.p.c. more than the arterial.

It was first shown by Pflüger (1864) that, if whole blood were very thoroughly evacuated, all the CO₂ could be removed without the addition of acid. On the other hand, all the CO₂ in plasma could not be removed by vacuum alone. After very thorough evacuation the addition of acid or of red blood corpuscles allowed the liberation of more CO₂. These facts suggested that, in plasma, CO₂ is present mainly as bicarbonate, for on

TABLE 32
Solubility of CO₂ in physiological fluids at
body temperature

FLUID	ABSORPTION COEFFICIENT ¹
Water	0.545
Plasma	0.510
Red cells	0.44
Whole blood	0.48

exposure to a vacuum NaHCO₃ loses a part of its CO₂ according to the equation,



To liberate the CO₂ in Na₂CO₃ it is necessary to add acid. These results suggested, too, that the red cells contain something which can act as an acid. We now know that the hemoglobin in the red cells is an amphoteric electrolyte and hence capable of acting as an acid.

The amount of dissolved CO₂ in blood can be calculated from its solubility coefficient in blood (table 32), when the pressure of carbon dioxide with which the blood is in equilibrium is known (p. 397). The dissolved CO₂ consists in part of carbonic acid (H₂CO₃). Although the actual amount of H₂CO₃ is extremely small (being only about 0.1 per cent of the dissolved CO₂) it is of great importance. When H₂CO₃ enters the blood, which is a slightly alkaline solution, it combines with base to form bicarbonate (BHCO₃) until an equilibrium is reached between the three forms of CO₂:



The relative amounts of these three forms at equilibrium depend upon the pH of the solution and can be calculated from the Henderson-Hasselbalch equation

$$\text{pH} = \text{pK}_1 + \log \frac{(\text{BHCO}_3)}{(\text{CO}_2 \text{ dissolved})^2}$$

pK₁ is a composite constant which in normal plasma has a value of about 6.1. The pH of plasma is normally about 7.4. Hence we may write

$$7.4 = 6.1 + \log \frac{\text{Bicarbonate CO}_2}{\text{Dissolved CO}_2^1}$$

or

$$\frac{\text{Bicarbonate CO}_2}{\text{Dissolved CO}_2} = \frac{20}{1}$$

¹ At a given temperature (H₂CO₃) is a constant fraction of the dissolved CO₂ and hence does not require separate representation in the formula.

² Cubic centimeters CO₂ (measured at STP) dissolved in 1 cc. fluid at a pressure of CO₂ of 760 mm Hg.

If the pH of the plasma is abnormal this ratio will, of course, be different.

The interior of the red cell is more acid than the plasma. Consequently, the ratio of bicarbonate to dissolved CO₂ will be smaller. Furthermore, since the water content of the red cell is less than that of plasma, the amount of dissolved CO₂ will be less too. For these two reasons then, at the same tension of CO₂ the cell contains less total carbon dioxide than the plasma. Yet, as we shall soon see, the red cells play the dominant rôle in the transport of carbon dioxide.

THE RÔLE OF HEMOGLOBIN IN CARBON DIOXIDE TRANSPORT

When carbon dioxide enters the blood from the tissues it combines with water to form H₂CO₃. This reaction is relatively slow in most solutions with a pH close to neutrality. In the blood, however, the reaction is catalyzed by an enzyme, carbonic anhydrase, which is found in the red cell but not in the plasma. The H₂CO₃ is thus formed within the cells. Nearly all of the H₂CO₃ thus formed then combines with base to form bicarbonate. The base available for combination is that which is already combined with weaker acids, mainly proteins, which are displaced by carbonic acid according to the equation,



Hemoglobin is used as the example in this equation because it does, in fact, furnish directly and indirectly,

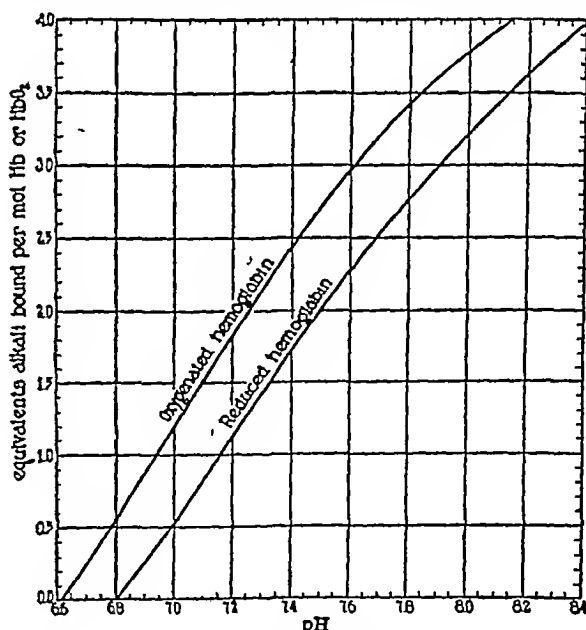


FIG. 32.7. Curves to show the amounts of base (K) bound by oxygenated and reduced hemoglobin at varying pH. The slopes of the curves represent the buffer values, in terms of the equivalent of base required to cause unit pH increase (from Peters and Van Slyke, *Quantitative Chemical Chemistry*, Vol. 1, 1932, from data of Hastings, Van Slyke, Neill, Heidelberger and Harrington).

the greater part of the base used. It does so for a number of reasons. In the first place, it comprises about three quarters of the total protein in blood. Secondly, it holds in combination an even greater proportion of the base held by weak acids in the blood, because it has so many weak acid groups in its molecule. Thirdly, it has the remarkably property of changing its acid strength with its degree of oxygenation. When the blood is in the tissues it loses its oxygen and the hemoglobin becomes a weaker acid and able to yield more base to carbonic acid. In the lungs the hemoglobin is oxygenated and becomes a stronger acid. This assists in displacing carbonic acid from combination with base and in turning it out of the blood. Figure 32.7 shows the titration curves of oxyhemoglobin and hemoglobin. It will be seen that at the same time pH oxyhemoglobin is combined with

more base than is hemoglobin, i.e., oxyhemoglobin is the stronger acid.

That hemoglobin behaved in this peculiar way was first suggested by Christiansen, Douglas and Haldane (1914), who investigated the carbon dioxide dissociation curve of reduced and oxygenated whole blood. As in the construction of an oxygen dissociation curve, samples of oxygenated or reduced whole blood are brought into equilibrium in a series of saturating vessels, called tonometers, with different pressures of CO_2 . The CO_2 contents of the equilibrated bloods are then determined by analysis and plotted against the corresponding gas tensions. Christiansen, Douglas and Haldane found that the curve for oxygenated blood was lower than that for reduced blood. In other words, reduced blood could carry more CO_2 at the same tension of CO_2 than oxygenated blood (fig. 32.8). For many years it was thought that this phenomenon was due entirely to the change in acid strength of hemoglobin on oxygenation. Reduced hemoglobin, being a weaker acid, would yield more base to carbonic acid and hence, at equal pressures of CO_2 , more bicarbonate would be formed. It now appears, however, that fifty per cent or more of this greater CO_2 -combining power of reduced blood is due to the greater power of reduced hemoglobin to combine directly with CO_2 (see p. 399).

Leaving aside the question of how the greater CO_2 -combining power of reduced blood is effected, let us consider the physiological importance of the phenomenon. By examining figure 32.8, it will be seen that if the blood, represented by point A on the curve, took up 5 v.p.c. of CO_2 , from the tissues and no reduction of the hemoglobin occurred, the tension of CO_2 in the blood would rise by about 14 mm Hg. If, however, about 6 v.p.c. of O_2 are lost from the capillary blood (as indicated by point V) the extra CO_2 can be taken on with a rise of only 7 mm Hg in the CO_2 tension. As the change in CO_2 tension is minimized, so too is the change in the pH of the plasma, because at the lower pressure of CO_2 less free carbonic acid is present. In the lungs the reverse reactions occur. Here oxygenation of the hemoglobin reduces the CO_2 -combining power of the blood, and a smaller fall in CO_2 pressure is effective in removing the excess CO_2 .

The dominant rôle of the red cells in CO_2 transport can be further demonstrated by contrasting the CO_2 dissociation curves of separated plasma and true plasma. The latter is constructed by exposing whole blood to different pressures of CO_2 and then separating the plasma and analyzing it for CO_2 , the former by equilibrating plasma in the absence of red cells (fig. 32.9).

It is apparent that the curve for separated plasma is much flatter than that of true plasma which signifies that much larger changes in CO_2 pressure and pH accompany a given change in CO_2 content in separated plasma. Evidently separated plasma is not as well buffered as true plasma. The greater buffer power of true plasma must be due to the red cells. We know,

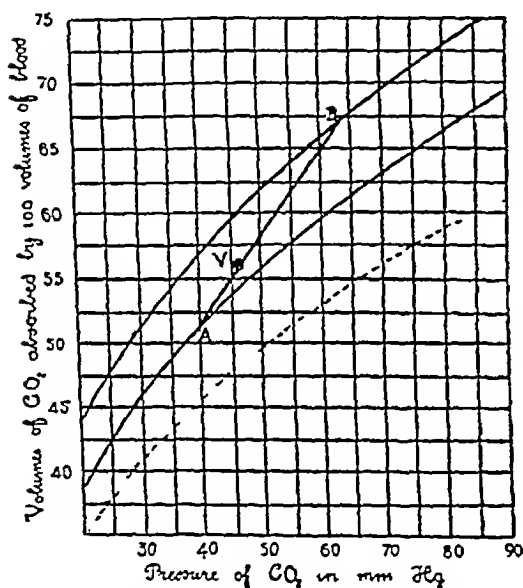


FIG. 32.8 Carbon dioxide dissociation curves of fully reduced human blood (upper solid line) in presence of hydrogen, and fully oxygenated human blood (lower solid line), in presence of air. Volumes of CO_2 along the ordinates, CO_2 tension along the abscissae. Line AVB is the so-called physiological dissociation curve of CO_2 , i.e., as a result of the reduction of hemoglobin the relation of volumes of CO_2 to CO_2 tension is indicated at points along this line and not along the lower curve for oxygenated blood. At A (arterial point) are indicated the volume and tension of CO_2 in arterial blood. Point B indicates the conditions in fully reduced blood. Point V (venous point) represents the degree of reduction of hemoglobin which normally occurs in the body during the passage of the blood through the systemic capillaries. The position of the line AVB varies with the respiratory quotient, moving to the right or left, respectively, with a rise or fall in the R.Q. Its position in the figure corresponds to a respiratory quotient of about 0.8. The interrupted line below is the CO_2 dissociation curve for oxygenated dog's blood. (Modified from Christiansen, Douglas and Haldane.)

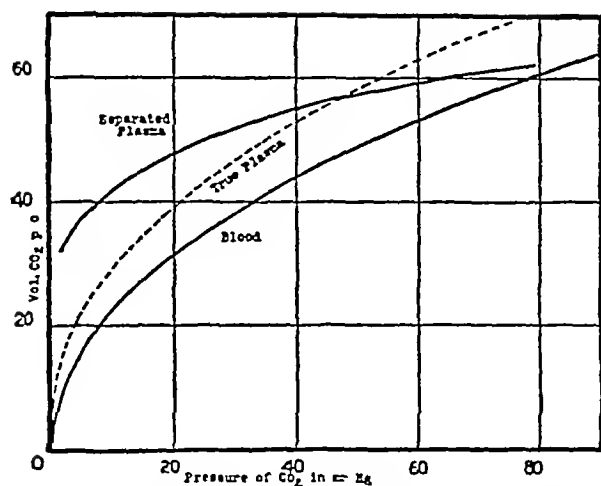


FIG 32.9 Comparison between the CO_2 dissociation curves of blood and of separated and true plasma from the same blood (After Evans, redrawn from data by Joffe and Poulton)

however, that base ions, e.g., K^+ and Na^+ cannot pass from the red cell into the plasma, because the red cell membrane is impermeable to cations with the exception of H^+ . Consequently, the red cells must increase the buffer power of the plasma indirectly. They do so by means of the *chloride shift* or *Hamburger phenomenon*.

If CO_2 is added to whole blood and the corpuscles and plasma are analyzed separately, it will be found that

- (1) The bicarbonate content of both plasma and corpuscles has increased
- (2) The chloride content of the corpuscles has increased
- (3) The chloride content of the plasma has decreased
- (4) The metallic cation content of the corpuscles and plasma has not changed
- (5) The water content and volume of the corpuscles has increased

The reverse changes occur when CO_2 is removed from blood by evacuation. Evidently, there is a transfer of the chloride from plasma to cells when CO_2 enters the blood, and the reverse process when CO_2 leaves the blood, hence the name, *chloride shift*. A simple qualitative explanation of this phenomenon can be given. When CO_2 enters the blood more HCO_3^- is formed in the corpuscles than in the plasma because they contain more available base for neutralizing H_2CO_3 . These excess HCO_3^- ions tend to diffuse out into the plasma, but, owing to the electrostatic attraction of the cations within the cells, can only do so if an equal number of Cl^- ions enter to take their place. Thus HCO_3^- from the cells enters the plasma in exchange for Cl^- which enters the corpuscles. This process will continue until an equilibrium is reached which has been found to agree (very nearly) with the distribution required by

Donnan's theory of membrane equilibria (p. 130). This requires the following relations³

$$\begin{aligned} \frac{[\text{H}^+]_{\text{cells}}}{[\text{H}^+]_{\text{plasma}}} \times \frac{[\text{Cl}^-]_{\text{cells}}}{[\text{Cl}^-]_{\text{plasma}}} &= \frac{[\text{H}^+]_{\text{plasma}}}{[\text{H}^+]_{\text{plasma}}} \times \frac{[\text{Cl}^-]_{\text{plasma}}}{[\text{Cl}^-]_{\text{plasma}}} \\ \frac{[\text{H}^+]_{\text{cells}}}{[\text{H}^+]_{\text{plasma}}} \times \frac{[\text{HCO}_3^-]_{\text{cells}}}{[\text{HCO}_3^-]_{\text{plasma}}} &= \frac{[\text{H}^+]_{\text{plasma}}}{[\text{H}^+]_{\text{plasma}}} \times \frac{[\text{HCO}_3^-]_{\text{plasma}}}{[\text{HCO}_3^-]_{\text{plasma}}} \end{aligned}$$

or

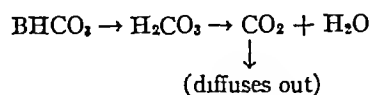
$$\frac{[\text{H}^+]_{\text{plasma}}}{[\text{H}^+]_{\text{cells}}} = \frac{[\text{Cl}^-]_{\text{cells}}}{[\text{Cl}^-]_{\text{plasma}}} = \frac{[\text{HCO}_3^-]_{\text{cells}}}{[\text{HCO}_3^-]_{\text{plasma}}}$$

The phenomenon can, in fact, be explained very precisely in terms of the Donnan theory. When CO_2 enters the blood the ratio $\frac{[\text{HCO}_3^-]_{\text{cells}}}{[\text{HCO}_3^-]_{\text{plasma}}}$ increases because more base is available in the cells. Similarly because the buffer power of plasma is less than that of the cells $\frac{[\text{H}^+]_{\text{plasma}}}{[\text{H}^+]_{\text{cells}}}$ increases. The Cl^- of the plasma must now

pass into the cells in exchange for HCO_3^- until the ratios are again equalized at a new level. Since the new ratio is a higher one, the number of osmotically active particles in the cells must now be higher than in the plasma. Consequently, water enters the corpuscles to equalize the osmotic pressures of the corpuscles and plasma and the volume of the corpuscles increases. If stasis of blood occurs during the withdrawal of blood from a vein, an abnormal amount of CO_2 may accumulate and the relative volumes of corpuscles and plasma may be appreciably altered. Hence stasis is undesirable when the blood is required for purposes where the relative volumes of corpuscles and plasma must be measured, e.g., in the estimation of blood volume.

THE EVOLUTION OF CARBON DIOXIDE IN THE LUNGS

The pressure of CO_2 in the alveoli is kept by respiratory activity at a lower level than it is in venous blood, hence CO_2 diffuses from the blood into the alveoli. This disturbs the equilibrium between the three forms of CO_2 and causes reactions to proceed in the direction indicated below



It has been known for a long time that the reaction $\text{H}_2\text{CO}_3 \rightleftharpoons \text{CO}_2 + \text{H}_2\text{O}$ is inherently a slow one. It is known too that the blood spends only about one second in the capillaries of the lung and about the same time in the capillaries of active tissues. The velocity constants of these reactions in solutions other than blood are known too, and it can be calculated that if blood had

³ In applying the Donnan equilibrium to the red cell the only cation which can be regarded as diffusible through the red cell membrane is the H^+ ion.

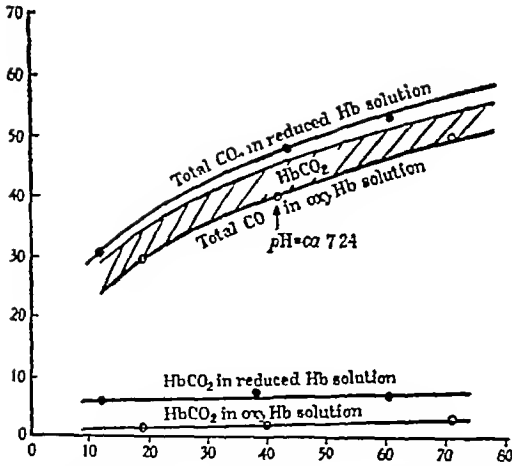


FIG 32 10 Upper curves, showing the proportions of the difference in CO₂ capacity of reduced and of oxy-hemoglobin which is due to the higher carbamino content of reduced hemoglobin. The shaded area represents the proportion due to carbamino CO. Lower curves, showing the negligible effect of a rise in CO₂ tension upon the carbhemoglobin content of blood (After Ferguson and Roughton, modified)

not certain peculiar properties, the time which it spends in the lungs would scarcely allow the escape of 10 per cent of the CO which we know does escape. Rapid loading and unloading of CO₂ by the blood is accomplished in two ways. The first, which has been mentioned already, is the action of the enzyme, carbonic anhydrase, which accelerates enormously, in either direction, the reversible reaction $\text{H}_2\text{CO}_3 \rightleftharpoons \text{CO}_2 + \text{H}_2\text{O}$. As the

carbonic anhydrase of the blood is found only in the red cells and not in the plasma, we may deduce that the rapid changes in CO₂ content occur primarily in the red cells while the plasma comes more slowly into equilibrium with the cells, perhaps after the blood has left the capillaries.

Carbonic anhydrase is present in a number of other animal tissues besides the erythrocytes, e.g., in gastric mucosa (p 499) and the mucosa of the small intestine, in the renal cortex (p 460), pancreas, lens and retina, brain, spleen, red muscle, liver, testes, in the oviducts of hens, where its function appears to be concerned in the deposition of calcium in the egg shell, and in saliva. Its physiological significance in most of these situations is unknown. Carbonic anhydrase is a protein, and zinc, as shown by Keilin and Mann, is an important constituent of its molecule. The action of the enzyme is destroyed by an inhibitor in plasma, which has been identified as a pseudoglobulin. It is also inhibited by sulfanilamide, cyanide, sulfocyanate, and by heavy metals. It is unlikely that the concentrations of sulfanilamide in the blood, such as are produced therapeutically, are sufficiently high to inhibit the enzyme. The site and mode of production of carbonic anhydrase are both unknown.

The other mechanism for the rapid combination and dissociation of CO₂ in blood is the direct combination of CO₂ with hemoglobin. This reaction does not go through the stage of carbonic acid and is very rapid. For many years past the existence of such a compound has been generally denied. Recent experiments, however, make it appear that about twenty-five per cent of the CO₂ liberated in the lungs under normal resting conditions has been carried in the blood in direct combina-

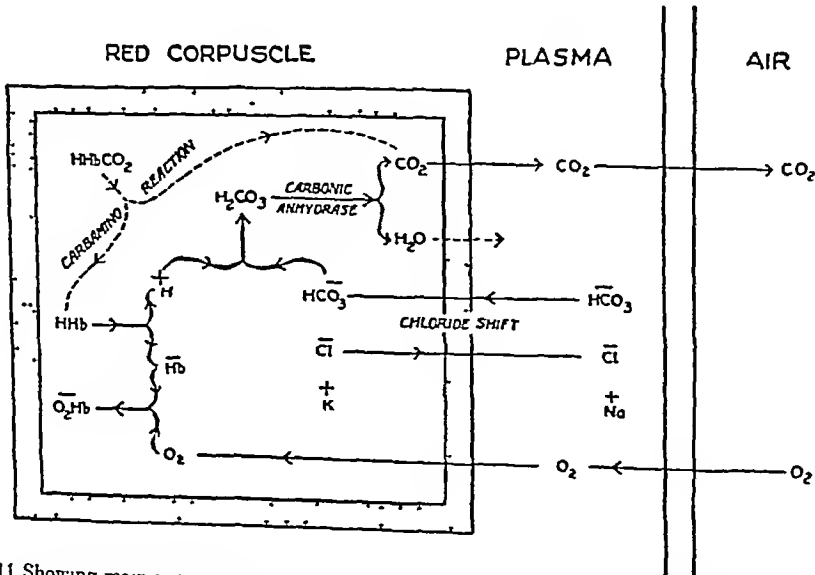


FIG 32 11 Showing main processes occurring in the blood during the output of carbon dioxide and uptake of oxygen in the lung (after Roughton)

tion with hemoglobin. Nevertheless, the total amount of CO_2 in the blood combined in this way is small, probably it never amounts to more than from 8 to 10 per cent of the total CO_2 . But the fact that reduced hemoglobin combines with more CO_2 than does oxyhemoglobin gives the compound an enhanced importance in the transport of CO_2 , it accounts for from 20 to 25 per cent of the gas freed in the lungs.

Carbon dioxide does not combine with the hemoglobin molecule in the same way that O_2 and CO do. It apparently combines with an NH_2 group to form a so-called carbamino acid. Hence, one name suggested for this compound is hemoglobino carbamic acid (Hb-NHCOOH). Another name less descriptive but possessing the virtue of brevity is carbhemo-globin.

Carbamino compounds of CO_2 with amino-acids are well-known and the technic of estimating these simpler compounds has been adapted for determining carbhemo-globin. Other forms of combined CO_2 , such as NaHCO_3 , can be precipitated as BaCO_3 by the addition of alkaline BaCl_2 . The barium salts of the carbamino-acids are soluble and remain in the supernatant fluid after centrifuging. The affinity of Hb for CO_2 diminishes with pH, and with strong acidification all the CO_2 dissociates off. Consequently, the Van Slyke technic for estimating total CO_2 can be applied to the supernatant fluid to measure the carbhemo-globin.

An increase in the CO_2 pressure of the blood should, *per se*, cause the formation of a greater amount of carbhemo-globin, but since an increase in CO_2 pressure is always accompanied by an increased acidity, which lowers the affinity of Hb for CO_2 , variations in CO_2 pressure over physiological ranges have actually little effect on the carbhemo-globin content of the blood. That is to say, the dissociation curve of carbhemo-globin is practically flat over physiological ranges of

CO_2 pressure. The main factor of physiological importance in displacing CO_2 from Hb is oxygenation of the Hb (fig. 32.10).

SUMMARY

About 5 per cent of the total CO_2 in blood is physically dissolved. Two to ten per cent, depending on the degree of oxygenation of the hemoglobin, is combined directly with hemoglobin (carbhemo-globin). The remainder is present as bicarbonate and, as such, is combined with base which has been yielded to H_2CO_3 by the weak acids of the blood, the most important of these is hemoglobin.

The CO_2 -combining power of reduced blood is greater than that of oxygenated blood, (1) because reduced hemoglobin is a weaker acid than oxyhemoglobin, and (2) because reduced hemoglobin can combine directly with more CO_2 to form carbhemo-globin than can oxygenated hemoglobin.

Base yielded by hemoglobin participates indirectly in the carriage of CO_2 by the plasma by means of the *chloride shift*. Base within the cells neutralizes the Cl^- ions which enter the red cells, thereby leaving base in the plasma free to neutralize HCO_3^- ions.

As the blood passes through the lungs it loses a small part of its total CO_2 (i.e., about 10 per cent). The elimination of CO_2 is accomplished with minimal change in pH and in CO_2 tension by the concurrent oxygenation of the blood which decreases the CO_2 -combining power of the blood in the two ways mentioned above.

The transfer of CO_2 , to and from the blood while it is in the capillaries, can be accomplished in less than one second, because (1) carbonic anhydrase catalyzes the slow reaction $\text{H}_2\text{CO}_3 \rightleftharpoons \text{CO}_2 + \text{H}_2\text{O}$ and (2) because the formation and dissociation of carbhemo-globin is inherently rapid (fig. 32.11).

CHAPTER 33

THE CONTROL OF RESPIRATION, PERIODIC RESPIRATION, DYSPNEA

THE LOCATION OF THE NERVOUS CENTERS OF THE RESPIRATORY MECHANISM. SPONTANEOUS ACTIVITY OF THE RESPIRATORY CENTER

It is customary and convenient to speak of the collections of nerve cells in the brain stem which discharge impulses to the muscles of respiration as the *respiratory center*. But the use of the term should not be taken to imply that the controlling nervous elements are in any sense a compact circumscribed mass or confined to a closely restricted area. Section through the brain at any level rostral to the upper border of the pons does not alter significantly the respiratory rhythm. But sections at various levels behind this cause pronounced disturbances in respiration. If a cut is made through the medulla, caudal to the tip of the calamus scriptorius, all breathing ceases. Those neurons that are distributed throughout the brain stem, and which integrate with the complex mechanism controlling the respiratory movements, will be here referred to as the respiratory center. The nerve cells are located at different levels in the formatio reticularis, and collected into functionally discrete units. The primary or functionally dominant part of this neural organization (and many restrict the term, respiratory center, to this part alone) lies in the lower third or so of the medulla (see p 401).

The classical experiments of Markwald (1887) form the basis of our knowledge of the location of the nervous structures controlling respiration.¹ He described powerful and prolonged tonic inspiratory movements or "cramps" which supervened after bilateral section of the vagus nerves and division of the brain stem immediately behind the posterior colliculi. He concluded that a center inhibitory to inspiration was located in the latter situation, but that the vagi also had an inhibitory action, consequently the inspiratory

"cramps" appeared only after vagal influence had been abolished as well. Markwald's observations were confirmed shortly afterwards by other experimenters. The subject was taken up in recent years by Lumsden, who found that the prolonged inspiratory movements occurred only if the section passed through the pons a few millimeters behind its anterior border and occurred whether or not the vagi were divided. The inspiratory cramps or *apneuses*, as Lumsden preferred to call them, last for several seconds. He postulated their dependence upon an *apneustic* or *inspiratory center* at the level of the striae acousticae, which was dominated normally by an inhibitory or *pneumolaxic center* situated in the upper part of the pons. The duty of the latter center was, through its inhibitory influence, to transmute the apneustic movements into the rhythmical movements characteristic of normal respiration. After section of the brain stem behind the striae acousticae the respirations consisted of a series of gasps occurring at relatively long intervals. Lumsden concluded that these represented the activity of a primitive *gasping center* situated in the lower part of the medulla from which the two higher centers had evolved. It was considered to be of little importance in higher animals.

Lumsden's results have been confirmed in the main by Stella and by Pitts, Magoun and Ranson. Stella, however, found in contradiction to Lumsden that section through the pons (i.e. separation of the pneumotaxic center) did not result in apneusis unless the vagal influence was abolished also (see fig 33 1). The results of Pitts and his associates are in essential agreement with those of Stella. They found that animals decerebrated through the upper part of the pons maintained a normal type of respiration which responded in the usual way to chemical and peripheral nerve stimulation so long as the vagi were intact, but immediately developed apneustic respiration and a complete cessation of rhythmical movements when both vagi were severed. Stimulation of the central end of one of the cut vagi temporarily restores the respiratory rhythm. The apneustic center is therefore under a double inhibitory influence, either one of

¹ Legallois (1824) located the respiratory center in the lower part of the medulla oblongata, and Flourens (1842, 1858) found a small spot about the size of a pin's head just beneath the forepart of the *calamus scriptorius* in the floor of the fourth ventricle, on either side of the mid line. He showed that after bilateral destruction of this area, which he named *noeud vital* (vital knot) the respirations ceased.

which is capable of converting the apneustic type of respiration to the rhythm of normal or nearly normal respiration. The vagal impulses influencing the apneustic center are initiated by the stretch of the lung towards the latter part of the inspiratory phase of normal breathing (see p 404)

The apneuses, like normal respirations, are affected powerfully by the CO_2 tension of the blood, being increased in depth by having the animal breathe an air mixture containing a high concentration of CO_2 and reduced in depth, or prevented from occurring, by carbon dioxide lack (see p 412). According to Stella, the pneumotaxic center is bilateral but its connections with the apneustic center are mainly homolateral, i.e., uncrossed

Pitts, Magoun and Ranson describe the respiratory center in the cat, which they locate in the reticular formation of the medulla, as consisting of an inspiratory and an expiratory division, both centers are bilaterally represented. The *inspiratory center* occupies the rostral half or two thirds of the reticular formation overlying the olivary nuclei on both sides, i.e., beneath the caudal third of the floor of the fourth ventricle (fig 33 2).² When stimulated a maximal co-ordinated inspiration results, involving both diaphragm and thorax. If stimulated during apneuses, the magnitude of the inspiratory movement is increased, if stimulated during an interval between apneuses an apneustic movement is produced. The *expiratory center* lies in the reticular formation dorsal to the inspiratory center. Electrical stimulation within this area causes expiration, if stimulated during inspiration or during apneusis these movements are inhibited. Regular respirations—inspiration alternating with expiration—are induced by rhythmical stimulation of the inspiratory center, expiration then occurs passively. Rhythmical stimulation of the expiratory center also produces regular respiration, spontaneous inspirations then alternating with the expiratory movements. Intimate synaptic connections exist between the diffusely distributed neurons within each center and between the two oppositely acting centers on the same side, as well as between similarly acting centers on opposite sides of the medulla. Thus, unilateral stimulation of a small area of the inspiratory center causes maximal contraction of all the inspiratory muscles. Excitation of one center causes simultaneous activation

² It is probable that the center is similarly located in man. Finley has described a case of respiratory failure in which a lesion was found post mortem in this situation

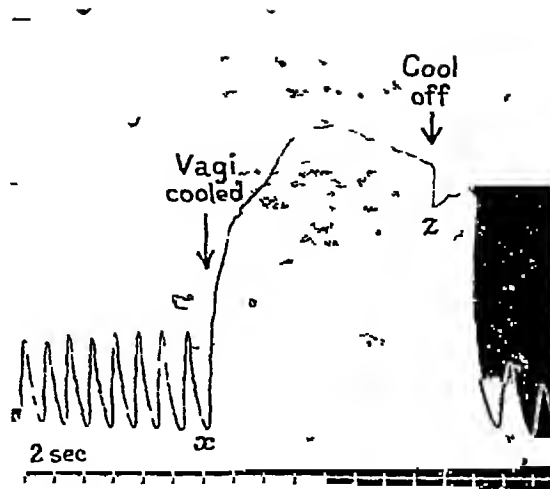


FIG 33 1 Showing apneusis. Section of the brain stem along a plane passing dorsally immediately behind the posterior colliculi, and ventrally 2.5 mm below the upper border of the pons. Between x and z the vagi were blocked by cold. Time 2 sec. (After Stella)

of its fellow of the contralateral side, accompanied by reciprocal inhibition of both oppositely acting centers. Thus, the alternating rhythm of inspiration is established and sustained, and the respiratory movements synchronized on the two sides of the body. The functional importance of the bilateral connections is demonstrated in a striking manner by making a deep longitudinal cut through the mid-line of the caudal part of the medulla,

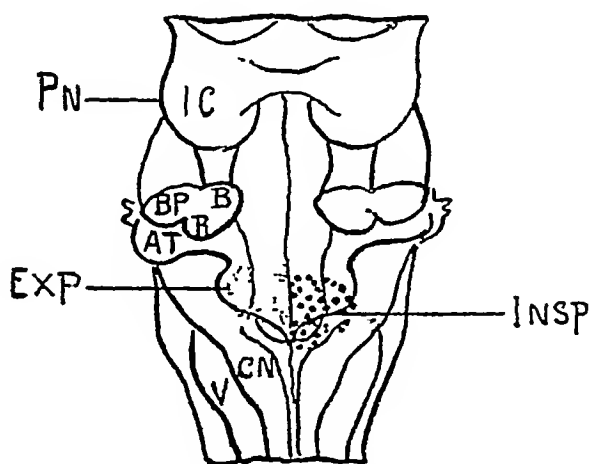


FIG 33 2 Dorsal view of lower brain stem of cat showing location of pneumotaxic (Pn), inspiratory (Insp) and expiratory (Exp) centers. Extent of expiratory center shown in light stippling, inspiratory center in heavy stippling. Though the centers are bilateral each is outlined on one side only, for the sake of simplicity. IC, inferior colliculus, AT, acoustic tubercle, B, brachium conjunctivum, BP, brachium pontis, R, restiform body, CN, cuneate nucleus (Redrawn from Pitts, Magoun and Ranson)

then the respiratory rhythms on the two sides of the body become asynchronous (see fig. 33.3)

The similarity in the effects of stimulation of the expiratory center and of the central end of the vagus, namely, inhibition of inspiration and of apneusis, has led Pitts and his colleagues to the conclusion that the vagal respiratory effects are mediated through the expiratory center. The influence of the pneumotaxic center is probably exerted in the same way.

The respiratory center is connected with the motor neurones of the phrenic and intercostal nerves in the cervical (C 3, 4 and 5) and upper thoracic segments of the cord (T 2-6) by descending tracts which run in the anterior columns and in the ventral part of the lateral columns of the spinal cord.

The spontaneous activity of the medullary respiratory centers has been a subject of interest to physiologists for many years. Some investigators, such as Coombs and Pile, and Schafer, have denied that the center is capable of spontaneous activity, main-

taining that afferent impulses, especially those set up in the lung by the stimulus of stretch and conveyed by the vagus, and those initiated from proprioceptors in the respiratory muscles and traveling by the posterior nerve roots, were essential. However, the results secured within recent years by means of improved methods of investigation leave little reason to doubt that, in certain species at least, the brain stem continues to discharge impulses to the respiratory muscles after all or nearly all afferent paths have been severed. For example, rhythmical bursts of impulses can be recorded from the central stump of the phrenic nerve of a decerebrate animal after section of the vagi, glossopharyngeal and other cranial nerves entering the pons and medulla, and division of the spinal cord below the level of the 7th cervical segment. Such an extensive operation would certainly interrupt all important afferent paths including those from the carotid sinus and the aortic arch. It is not to be supposed, of course, that the respirations would be normal after such a radical procedure, for even if not essential for maintaining the activity of the center, afferent nerve influences are of the utmost importance in the regulation of that activity and the production of the normal respiratory rhythm. Evidence for automaticity of the respiratory center of a cold blooded species has been secured by Adrian and Buytenjck. They succeeded in recording rhythmical action potentials originating in the vagal lobes of the excised brain stem of the gold-fish, the potential changes had the same range of frequency as the respiratory movements (fig. 33.4).

The spontaneous respiratory activity is apparently dependent primarily upon the inspiratory center, the expiratory center playing an inhibitory rôle to interrupt intermittently the inspiratory discharge. There is no evidence that impulses are discharged spontaneously from the pneumotaxic center. Pitts and his colleagues suggest that the pneumotaxic center is excited from the inspiratory center, that a proportion of the impulses discharged from the latter region ascend to the pontine center whence impulses descend to the expiratory center, a discharge of *inhibitory* impulses is transmitted from the latter to the inspiratory center. As the discharge of impulses from the expiratory center ceases, the inspiratory center resumes its activity and the cycle is repeated. A rise in the rate of discharge of the inhibitory impulses will increase the rate of breathing, while a reduction in the fre-

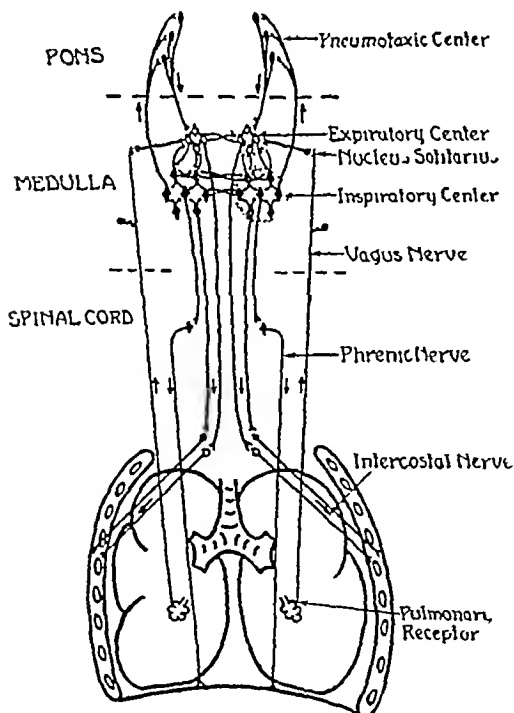


FIG. 33.3 Diagram illustrating the chief nervous connections responsible for the control of the respiratory rhythm. Inspiratory and expiratory centers in the medulla on the right side surrounded by dotted circles (After Pitts, slightly modified.)

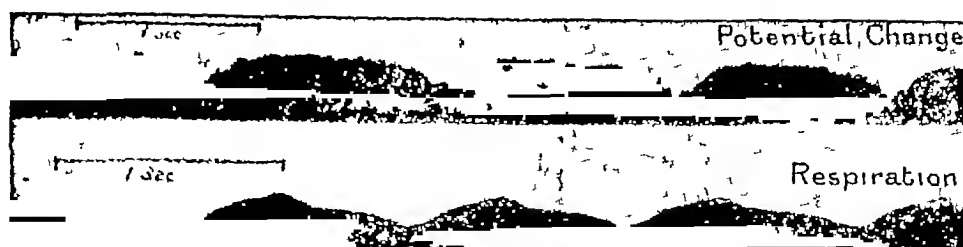


FIG 33-4 Comparison of electrical potential wave rhythm and rhythm of respiratory movements (After Adrian and Buytendijk)

quency of the discharge will be followed by slowing of the respirations

It appears, then, that in the intact animal the automatic alternating rhythm of the medullary centers is maintained by inhibitory impulses from the pneumotaxic and vagal centers impinging upon the expiratory centers, which in turn inhibit, intermittently, the activity of the inspiratory centers. The vagal influence is evoked by inflation of the lungs (i.e., by inspiration), expiration, a passive act, follows (p. 453)

Under ordinary physiological conditions the vagal regulating influence appears to play the dominant role. The respirations are immediately slowed and deepened when these nerves are cut. On the other hand, the rapid breathing caused by a rise in body temperature is affected only slightly by vagal section; the polypnea is then thought to depend mainly upon the activity of the pontine center as influenced by impulses from the hypothalamus. Injury to it or its separation from the medullary centers at once reduces or abolishes the rapid breathing. The reaction of the pneumotaxic center to a rise in temperature is probably brought about through its connections with the thermoregulatory centers in the hypothalamus.

The grading of the strength of the contractions of the inspiratory muscles is apparently brought about by (a) alterations in frequency of the twitches of the component muscle units and (b) recruitment or derecruitment of the individual muscle units. That is, more or fewer muscle units are brought into play.

Theories relating to the automatic (repetitive) discharge of impulses from the respiratory neurons

The frequency of the periodic discharge of impulses from the respiratory center, as well as the rate of the alternating rhythm of activity and inactivity displayed by the respiratory center, is influenced by several factors, e.g., carbon dioxide and oxygen tensions of the surrounding medium,

H-ion concentration, and the nature and frequency of the afferent impulses reaching the nerve cells. An *electronic theory* has been offered by Gesell to explain the repetitive discharge and the changes in its frequency under various conditions. A longitudinal polarization of the nerve cell is conceived; a potential difference is postulated between the dendrites and the point of origin (axon hillock) of the axon, electric currents flowing within the cytoplasm from the former processes to the latter and in the opposite direction extracellularly. The cell is stimulated presumably by the emergence of the currents at the axon hillock, thus producing the rhythmic discharge of impulses along the nerve fiber. The strength of these currents, which determine the frequency of the discharge, is conceived as varying with the chemical constitution of the cell's surroundings and with the frequency and character of the afferent impulses (especially from the carotid sinus and carotid body) bombarding its surface. This conception has been elaborated into an acid-humoro-electronic theory based upon the electronic action of acetylcholine. The magnitude of the potential differences are thought to depend upon the quantity of the neurohumor present at any instant at the synapses; this in turn depends upon the balance between the rate of production of acetylcholine and its destruction by cholinesterase. Carbon dioxide is postulated as acting to inhibit the destruction of cholinesterase. Among the experimental observations which Gesell and his associates advance in support of this theory are the actions of dipropylfluorophosphate and prostigmine upon respiration when injected into the vertebral artery. These drugs enhance cholinergic effects (chap. 72). The first mentioned was observed by Frey and Gesell to cause a slowly increasing hyperpnea and, later, strongly active movements of the accessory muscles of respiration. Stimulation of the vagus or sinus nerve of the poisoned animals resulted in strong after discharges of the respiratory responses.

Pitts has formulated a theory to account for the repetitive discharge of the medullary centers, based upon that advanced by Adrian and Zotterman to explain the impulse discharge of peripheral receptor organs in response to continuous stimulation. It may be stated briefly as follows. The rate of the impulse discharge at any moment depends upon the sum of all the excitatory and inhibitory effects, chemical and nervous, acting upon the respiratory neurons. The action of these corresponds to the constant stimulus in Adrian and Zotterman experiments. The passage of each impulse by such stimulation is followed by an absolute refractory period of very brief duration (2 milliseconds), and this in turn by a relative refractory period of much longer duration (100 milliseconds), in which excitability recovers progressively, i.e., a stronger stimulus is required to excite the nerve if applied early in this period than if applied late. The respiratory neuron, after the passage of the impulse, will recover its excitability and respond earlier in the relative refractory period if the constantly acting stimulus is strong than if it is weak. Thus, with a constantly acting stimulus a response occurs in, say, 25 milliseconds, whereas with a very weak stimulus a response will not occur until the lapse of 100 milliseconds. Thus, the interval between stimuli will be shortened by the stronger stimulation, i.e., a higher rate of repetitive discharge will be induced. As mentioned above, the repetitive discharge is converted by inhibitory impulses into the rhythm of alternating periods of activity and quiescence characteristic of the inspiratory center.

THE FACTORS REGULATING RESPIRATION

The factors affecting the activity of the respiratory center and, therefore, the volume of respired air (pulmonary ventilation) are classed for purposes of description and analysis into four groups, (a) voluntary and emotional influences, (b) reflex, (c) chemoreflex, and (d) chemical factors (direct action on respiratory center).

Voluntary control Emotional influences

That the respirations are under voluntary or semi-voluntary control for short periods of time is common knowledge. This control, though we are scarcely conscious of it in most instances, is being exerted in numerous ways in the ordinary affairs of daily life, such as in speaking, swallowing (p 559), laughing, blowing, coughing, sucking, etc.

But the power of the will to inhibit respiration is strictly limited. The breath can be held for only a brief space (a maximum of about 90 seconds) before automatic or involuntary control asserts itself, the inhibitory influence is over-ridden and the muscles of respiration contract despite all one's efforts to "hold the breath" (see also p 412). The nerve elements giving origin to the voluntary impulses are probably in the motor area of the cerebral cortex.

The respirations may be effected profoundly by impulses arising in higher cerebral centers as a result of various emotional or other mental states, e.g., fear, grief, surprise, interest, amusement, etc.

Respiratory effects are produced by stimulation of the orbitofrontal part of the cerebral cortex, stimulation of the anterior part of the cingulate gyrus (area 24, see ch 68) causes respiratory arrest in expiration.

Reflex control

Hering Breuer reflexes. The importance of afferent impulses from the lungs in the control of respiration was first pointed out in 1868 by Hering and Breuer, who showed that inflation of the lungs arrested inspiration, expiration then ensuing, while deflation inhibited expiration and brought on inspiration. These are reflex effects, mediated through the afferent fibers of the pulmonary vagi, for they are abolished after these nerves have been divided. In ordinary breathing the inflation reflex alone comes into play, thus checking, at the end of an inspiration of the usual duration, further distension of the lungs. When the inhibitory vagal influence is abolished by cutting the nerves the respiratory rate, as might be expected, is slowed, and the depth of breathing is increased, for the inspiratory phase then continues for a longer period before it gives place to expiration. The total pulmonary ventilation is little altered. As mentioned on page 400 inspiratory movements of such length as to abolish rhythmical respiration follow if the inhibitory influence emanating from the pontine (pneumotaxic) center is excluded as well. The corresponding deflation reflex does not apparently play a part under ordinary conditions (Adrian, Partridge). In order to elicit it, extreme deflation of the lungs is required, as can be produced in the laboratory by forcible compression of the chest, by collapse of the lungs through the production of a pneumothorax, or by sucking air from the trachea. As a result of these reflexes the respiratory

excursions in ordinary breathing are only a small fraction of what is possible in maximal hyperpnea.

Head, in a later study of these reflexes in Hering's laboratory, isolated a slip of the diaphragm and recorded its action during distension and deflation, respectively, of the lungs. When the trachea was clamped towards the end of a normal inspiration the muscle slip immediately relaxed and the inspiratory movement ceased. Blocking the trachea at the end of expiration called forth a powerful inspiratory movement, sucking air from the lungs caused a tonic contraction of the diaphragm.

Adrian approached the problem by recording during the phases of respiration the action potentials passing up a single afferent vagal fiber. The frequency of the impulses was found to vary with the degree of stretch of the lung, being highest when lung inflation was greatest (i.e., at the end of the inspiratory phase) and lowest when deflation was more nearly complete (i.e., towards the end of expiration) (fig. 33.5). The receptors which, as intimated above, are sensitive to a stretching force, are thought to be situated in the walls of the alveolar ducts, the most distensible parts of the lung structure. They adapt slowly (ch. 63), for a stream of impulses continues to ascend the nerve with relatively little reduction in frequency while inflation of the lung is maintained.

There are two interesting observations in connection with the effect of vagus nerve impulses on the rhythm of respiration, which may be mentioned here. (1) It was reported by Stewart, Pyke and Guthrie that the rhythm of respiration in a case which had been resuscitated from brain anemia was the same as that of the

artificial respiration which was given during the resuscitation process. (2) It is easy to show, when phrenic impulses (fig. 33.6) are recorded by a valve amplifier in a curarized animal, that the rhythm of the respiratory center comes into phase with that of the artificial respiration. The phrenic impulses coincide with the inspiratory phase of artificial respiration as long as the vagi are intact. This correlation is lost when the vagi are cut.

Chemoreflex control

Heymans and Heymans made the surprising discovery in 1927 that respiratory reflexes could be elicited from the aortic area. This observation and the brilliant researches of Heymans and his associates, in the years following, on the corresponding rôle played by the structures in the region of the carotid bifurcation opened a new field in the physiology of respiration, and a fresh outlook upon the vexed question of respiratory control. Many preconceptions and misconceptions have been swept aside. A more detailed description of these areas has been given in chapter 27 where their functions in the control of the vascular system were discussed.

The carotid and aortic areas each contain two types of receptors, one type (pressoreceptors) responds to mechanical, the other (chemoreceptors) to chemical stimulation. The pressoreceptors situated among the collagenous fibers in the wall of the carotid sinus and in the wall of the aortic arch, are stimulated by a stretching force, as by a rise in arterial blood pressure. The chemoreceptors are contained in small gland-like structures—the *carotid* and *aortic bodies* (p. 283). The respiratory

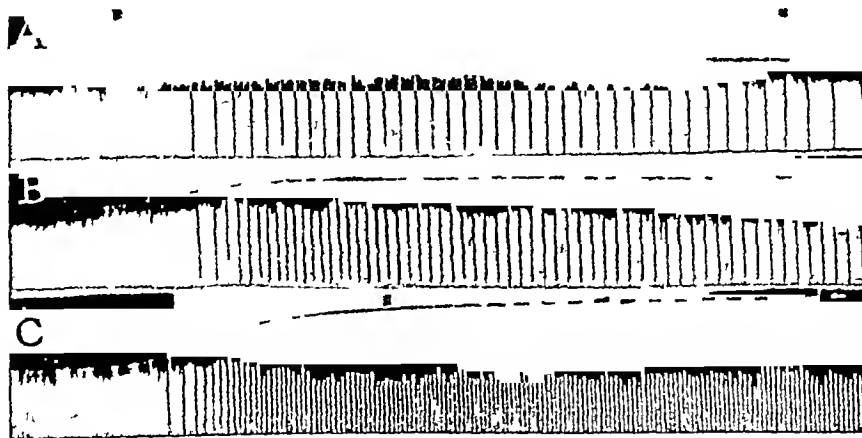


FIG. 33.5 Records of impulses discharged over vagus nerve during inflation of lung. Spinal cat, single fiber preparation. Inflation of the lungs by pump. Movement of signal line directly proportional to inflation. A, inflation = 65 cc, maximum frequency, 80 per second; B, inflation = 115 cc, maximum frequency, 120 per second; C, inflation = 230 cc, maximum frequency, 250 per second. (After Adrian.)

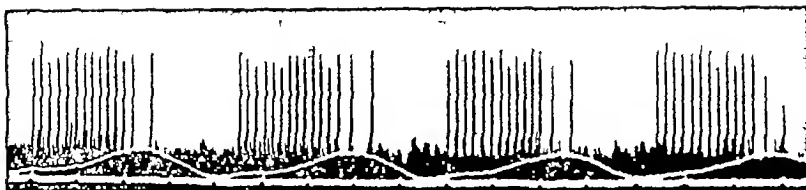


FIG 33 6 Showing discharge of impulses over a single fiber of the phrenic nerve during artificial respiration (wavy line) in a curarized animal (After Partridge)

reflexes initiated from these two types of receptor are contrary in their effects. Stimulation of the pressoreceptors inhibits respiration, an abrupt rise in blood pressure such as follows the injection of adrenaline causing respiratory arrest (apnea). Excitation of the chemoreceptors increases the rate and depth of breathing. Both types of reflex are abolished by section of the supplying nerves (sinus or aortic).

Though of the utmost importance in the control of the circulation, the *pressoreceptors* do not appear, in mammals at least, to serve any respiratory function under physiological conditions. With the possible exception of their functioning in very strenuous exercise (p 411) there is no circumstance of a physiological nature under which the inhibition of respiration is caused by a rise in blood pressure or by the stimulation of these receptors in any other way. Nor is it apparent that such a reaction would serve any useful purpose. That they give a respiratory response when very strongly stimulated appears to be accidental and without physiological significance. They may possibly represent the vestige of a mechanism which was of value in some aquatic ancestral form and to which the mammalian body has fallen heir.

The *chemoreceptors* are stimulated by oxygen lack, but not until the oxygen tension of the arterial blood reaches a relatively low level. A man breathing an atmosphere of which the oxygen percentage is gradually lowered shows little or no change in respiration until the oxygen is reduced to less than 16 per cent, and a reduction even to about 10 per cent causes only a moderate increase (17 per cent) in pulmonary ventilation. The hyperpnea is caused by stimulation of the respiratory center through chemoreflexes originating in the carotid and aortic bodies. They constitute the sole protection against anoxia, for we shall see that the effect of oxygen lack directly upon the center is one of depression.

The chemoreceptors are less sensitive still to

carbon dioxide. When, for example, the carotid body is isolated from the general blood stream (effect upon the respiratory center thus eliminated), but its nerve supply retained, and then perfused with a solution containing CO_2 , the smallest change in CO_2 tension which causes an appreciable change in respiration is around 10 mm Hg. When, on the other hand, the reflex effect is abolished by denervation of the carotid area, a change in CO_2 tension of only 3 mm Hg in the blood supplying the respiratory center is sufficient to induce hyperpnea (Schmidt, Dumke and Driggs).

Schmidt and his associates conclude from their experiments that the chemoreceptors do not play a rôle in the control of respiration under ordinary physiological conditions. This conclusion receives support from the observations of Cromer and Ivy, who found that dogs worked just as well on a treadmill after as before excision of the carotid bodies. Also, according to Dautrebande, a dog with its carotid and aortic chemoreceptors removed carries out respiratory adjustments to high altitudes as well as a normal animal.

In more exacting emergencies, however, the reflex response of the chemoreceptors, especially to anoxia, is undoubtedly of the highest importance. Anoxia appears to be ineffective as a stimulus to the respiratory center. Depression and ultimate failure of the central neurons is the predominant effect of oxygen lack. The integrity of the chemoreflex mechanism, on the other hand, according to Comroe and Schmidt, is highly resistant to anoxia, retaining its viability and continuing to exert its influence upon the center which otherwise would be unresponsive in the body's emergency. These observers look upon the chemoreceptor mechanism as a more primitive type of respiratory control which serves as a last line of defense against respiratory failure—the *ultimum moriens* of the respiratory control system.

The chemoreceptors also appear to be relatively

insensitive to the H ion. Such slight variations in hydrogen ion concentration of the blood as might occur under any condition of health could have little or no stimulating effect upon them. In experimental animals a relatively large shift in pH (namely 0.1) of the fluid perfusing the isolated carotid body is required in order to effect a change in respiration, it exerts a somewhat greater effect directly upon the center itself.

The experiments of Watt, Dumke and Comroe provide evidence for tonic chemoreflex activity. They found that the inhalation of pure oxygen by unanesthetized dogs caused a reduction in breathing of as much as 31 per cent, an effect which was abolished by the denervation of the carotid and aortic bodies. This indicates that while breathing ordinary air some at least of the chemoreceptors had been tonically active. The authors concluded, nevertheless, that such activity was of little significance in the regulation of normal breathing for the reduction in pulmonary ventilation was transient, the breathing in most instances returning to its original level within 3 to 6 minutes though pure oxygen was still being breathed. Furthermore, denervation of the chemoreceptors (without oxygen inhalation) caused no change in breathing. In man, *under anesthesia*, in *congestive heart failure*, *pulmonary emphysema*, etc., the breathing of pure oxygen may, however, cause complete cessation of breathing for a time. This "oxygen apnea" is attributed to the removal of the stimulating effect of anoxia upon the peripheral chemoreceptors while the respiratory center is depressed to the point of inactivity by the anesthetic or the anoxia itself (see Oxygen Therapy, ch. 34).

REFLEXES INITIATED IN OTHER PARTS OF THE BODY

The carotid and aortic bodies are probably not the only structures containing chemoreceptors, through stimulation of which reflex respiratory effects can be evoked. A small collection of tissue, structurally similar to the carotid and aortic bodies, has been discovered in the vicinity of the pulmonary artery, and Pi-Suner has shown that hyperpnea results from asphyxia after inactivation of the carotid and aortic chemoreceptors. This he attributes to reflex effects initiated from chemoreceptors in the lungs or tissues, or in both these situations. The experiments of Churchill and Cope also show that reflex increase in the respiratory rate results from distention of the pulmonary vas-

cular bed. When the vessels of the lung, isolated from the body except for its nerves, was injected with fluid to cause overfilling of the vascular bed, rapid shallow breathing immediately ensued. Section of the vagal fibers abolished the response. Weak respiratory responses to anoxia can be elicited from abdominal chemoreceptors.

Stimulation of almost any afferent nerve may bring about a reflex change in respiration. Stimulation of pain fibers is especially potent in this regard and the respiratory effects of the excitation of the cutaneous nerves by extremes of heat or cold are well known. The thermal effect upon respiration is apparent in the panting of the dog. The increased pulmonary ventilation in fever is also partly due to the stimulation of receptors (*thermoreceptors*), responsive to a rise in temperature. The thermoreceptors are situated peripherally, especially in the skin, and centrally in the hypothalamus, the latter upon being stimulated through the rise in temperature of the blood sets up impulses which are transmitted to the respiratory centers, the pneumotaxic center in particular. The great increase in pulmonary ventilation occurring in muscular exercise is dependent in part upon reflexes originating in the active muscles and moving joints. Proprioceptive impulses from the diaphragm and other respiratory muscles during one respiratory phase exert an important influence upon the succeeding movement.³ Stimulation of sensory nerves in the respiratory tract, as by ether anesthesia, or of the abdominal viscera, either during surgical operations or as the result of disease may cause profound changes in breathing. Also, as pointed out by Harrison and his associates, stimulation of afferent nerve endings in the great veins and right auricle by the rise in venous pressure is a factor in the hyperpnea of exercise or in the dyspnea of cardiac failure. The glossopharyngeal nerve contains afferent fibers which inhibit respiration during the second stage of the act of swallowing. Abrupt inhibition of respiration is also caused by the inhalation of an irritant gas.

³ Larsell has described specialized sensory endings in the lung tissue which possibly are chemoreceptor in function. The lung tissue would be an even more advantageous site for sampling the oxygen content of the blood than is the carotid or aortic arch. Pi-Suner's experiments suggest that a chemoreflex mechanism is present in this situation, but his results have so far not been confirmed by other workers. As Gesell has pointed out, the existence of such a chemoreceptor organ would explain the rapid breathing caused by multiple pulmonary emboli (p. 427) and by pulmonary edema.

through stimulation of nasal branches of the 5th nerve. In other instances irritation of these endings may cause sneezing—a modified respiratory act (see fig 33 7). Coughing, though it can be brought about by a voluntary effort, is most commonly reflex in character, initiated by the stimulation of afferent nerve endings in the trachea and larynx (p 355).

The pulmonary ventilation is increased to the greatest possible degree by voluntary hyperpnea, and may reach a value of 160 liters per minute. The reason for the volume of air breathed being less during muscular exercise than when the hyperpnea is voluntary is possibly due to the rise in blood pressure, which occurs in exercise and the consequent reflex inhibitory effect upon respiration referable to the carotid sinus mechanism.

CHEMICAL CONTROL THROUGH THE RESPIRATORY CENTER

HISTORY OF THE DEVELOPMENT OF MODERN VIEWS ON THE CHEMICAL REGULATION OF RESPIRATION UP TO THE DISCOVERY OF THE FUNCTIONS OF THE CAROTID AND AORTIC BODIES (1929)

The earliest definite chemical theory of respiratory control was proposed by Rosenthal, 1882, he believed that the oxygen tension of the arterial blood was the principal regulating factor. This view was based upon

the observations (1) that the respiration of animals breathing an atmosphere poor in oxygen was sharply increased, and (2) that during the cessation of respiration (apnea) caused by a short period of forced ventilation of the lungs, the oxygen content of the arterial blood was raised slightly. The theory, though attractive in the sense that it postulated a mechanism by which the want of essential oxygen was automatically corrected, but it failed to explain why breathing pure oxygen did not cause cessation of breathing. Furthermore, the hyperpnea caused by inhaling a high concentration of CO_2 , or that due to acidosis, is not accompanied by a reduction in the oxygen tension of the blood, and even in severe exercise, in which the volume of air breathed may be increased twenty times or more, no change occurs in the arterial oxygen tension (see table 33, p 410).

In 1885, Miescher-Rüsch called attention to the importance of the carbon dioxide tension of the blood as a controlling factor in respiration. This theory of the essential role played by CO_2 in respiratory control was later (1905) elaborated by Haldane and Priestley. In the ensuing years numerous experiments, especially upon man, were carried out by Haldane and his associates which established the fundamental importance of carbon dioxide in the regulation of the pulmonary ventilation. Originally, Haldane believed carbon dioxide to be the sole or at least the main chemical stimulant, and that it acted in a unique or specific manner. But in 1911 Winterstein suggested that the H-ion concentration of

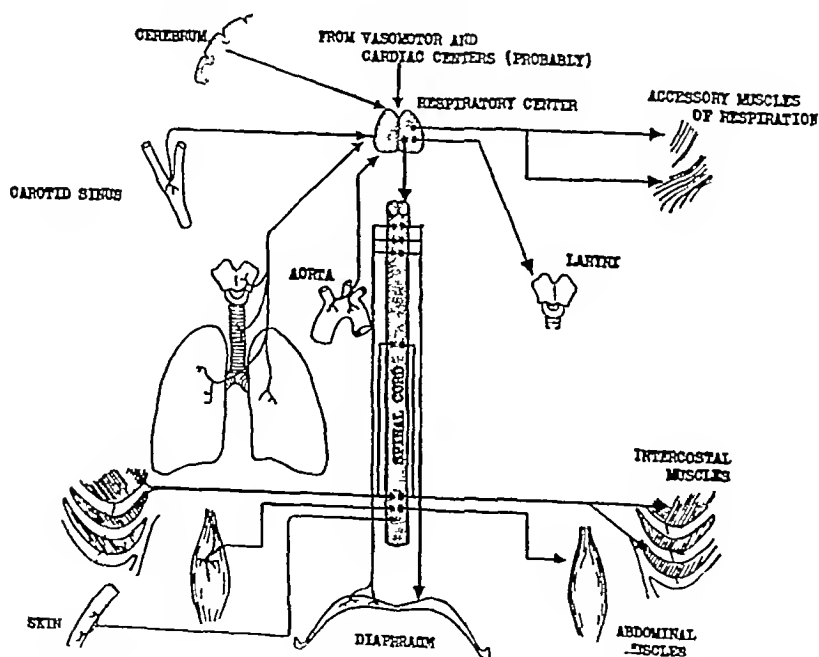
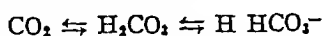


FIG 33 7 Diagram to illustrate the reflex and chemoreflex control of respiration

the arterial blood was the controlling factor, a view which was later supported by experiments and the work of Hasselbalch (1912) and his associates. They showed that any tendency for the blood pH to fall was counteracted by increased pulmonary ventilation and the elimination of CO_2 . The CO_2 tension of the alveolar air (and arterial blood) fell but the blood pH remained unchanged. It was then thought that the effect of CO_2 was due simply to its acting as an acid. It was soon found, however, that there was not always correlation between the pH of the blood and the pulmonary ventilation, a rise in pH might be associated with an increase in breathing, as in anoxia, hemorrhage, or injections of sodium bicarbonate. Experimenting with decerebrate cats inhaling CO_2 with and without previous slow injections of bicarbonate, Scott (1918) found that the hyperpnea caused by the inhalation was the same in both instances, i.e., irrespective of the pH of the blood. He concluded that a true alkalosis causes apnea only when, at the same time, there is a sudden fall in CO_2 tension, and that undissociated CO_2 , quite apart from its acid properties, acts as a specific respiratory hormone. Dale and Evans caused a rise in pH by bicarbonate injections, but this did not induce cessation of breathing as the H-ion theory demanded. Collip (1920) found that injections of bicarbonate actually stimulated breathing and concluded that the HCO_3 ion was the controlling factor, and not CO_2 or undissociated carbonic acid.⁴ Gesell (1925) in an effort to compose the discrepancies of previous theories, proposed that the determining factor was the H-ion concentration of the cells of the respiratory center itself, affected largely by the products of its own metabolism, and not necessarily that of the arterial blood. The ready penetration of CO_2 across cell membranes was given in explanation of the much greater stimulating effect of CO_2 upon respiration than that of mineral acids. Jacobs (1920) had found that undissociated carbonic acid penetrated living cells more readily than did any other acid of an equivalent hydrogen ion concentration, the acidity of the interior of living cells immersed in an alkaline solution of CO_2 and sodium bicarbonate was raised to nearly the same degree as when placed in an acid solution composed of CO_2 in distilled water. The ready diffusion of CO_2 across cell membranes accounted for its greater lethal effect upon tadpoles, and its more intense stimulation of human taste buds as compared with mineral acids.

In 1929 the sensational discovery of the functions of the peripheral chemoreceptors (carotid and aortic bod-

⁴ Carbon dioxide in aqueous solution, is present in three forms, dissolved CO_2 , undissociated carbonic and dissociated carbonic acid. Thus



Adding sodium bicarbonate to the solution causes a shift of the reaction to the left, which increases the free CO_2 and undissociated carbonic acid at the expense of the dissociated. Weak, undissociated acids were shown by Loeb to diffuse freely through cell membranes

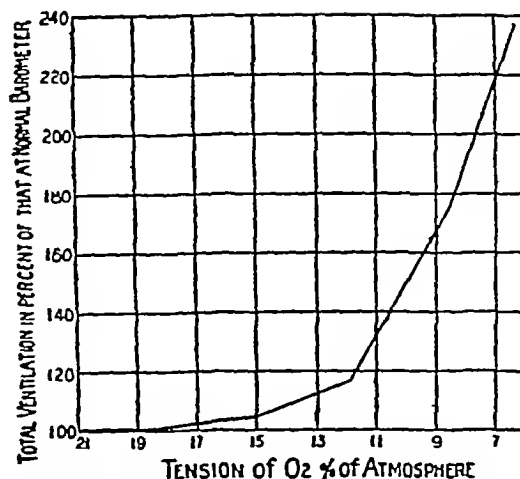


FIG. 33.8 Effect of oxygen lack upon pulmonary ventilation in human subject. (After Means)

ies) in respiratory control was reported by Heymans. Several discrepancies were explained in the light of this discovery. The part which these structures play in the regulation of pulmonary ventilation has been dealt with. In 1936 Nielson secured evidence in support of Haldane's original conception of the specific nature of CO_2 in the control of breathing, and relegated other chemical factors to the mere role of sensitizers of the respiratory center. He found, for example, that in human subjects, though the hyperpnea caused by the inhalation of CO_2 was extreme, little change in arterial pH resulted. The ingestion of ammonium chloride, on the contrary, while producing a much more pronounced lowering of the pH of the arterial blood caused little or no increase in breathing. It would seem that the question of the greater penetrating power of carbon dioxide can be ignored in considering these results, for the experiments in which ammonium chloride was ingested were continued over a period more than long enough (up to 10 days) for the cells of the center to have come into equilibrium with the arterial blood.

MODERN VIEWS ON DIRECT CHEMICAL EFFECTS ON THE RESPIRATORY CENTER. The principal chemical factors which influence the activity of the respiratory center are the CO_2 and O_2 tensions (pCO_2 and pO_2) of the arterial blood and the H-ion concentration. Receptors (chemoreceptors) capable of stimulation by pCO_2 or the H-ion are present in the inspiratory center. A rise in the pO_2 tension of the blood above normal does not influence the activity of the center. But a fall in oxygen tension (anoxia or hypoxia), if considerable, depresses the center and if the reduction in pO_2 is severe, the center becomes unresponsive to ordinary chemical stimulation, its activity then being maintained only through impulses from the peripheral chemoreceptors (carotid and aortic bodies) which are

much more resistant to anoxia. Thus, the hyperventilation caused by anoxia is entirely through the stimulation of peripheral chemoreceptors.

The effect of $p\text{CO}_2$ upon the respiratory center
Under the ordinary physiological conditions of the body at rest or engaged in mild or moderate exercise, the carbon dioxide tension of the arterial blood is the principle factor, through its central action, in the regulation of respiration. Through its effect in increasing the pulmonary ventilation, its percentage in the alveolar air (and therefore its tension in the arterial blood) is maintained remarkably constant at around 5.5, though the percentage in the inspired air is varied very widely. Thus in an experiment of Haldane and Priestley, the percentage of carbon dioxide in the air breathed was increased to nearly 4 per cent (normal 0.04 per cent), the percentage in the alveolar air rose from 5.62 to 5.97, an increase of only 0.35 per cent. Yet this small rise, equivalent to an increase in partial pressure of between 2 and 3 mm Hg, increased the pulmonary ventilation over 2½ times. An increase in CO_2 percentage much less than this (a twentieth) causes an appreciable increase in the volume of respired air. A reduction in alveolar CO_2 percentage acts in the opposite manner, a fall of 0.2 per cent (equivalent of 1.5 mm Hg $p\text{CO}_2$), as can be readily induced by forced breathing, causes apnea (p. 412). It is not, however, so much toward the maintenance of a constant percentage of CO_2 in the alveolar air that regulation is directed as toward maintaining the normal alveolar CO_2 tension. This becomes apparent when a subject is exposed in a chamber to a pressure of several at-

mospheres. The CO_2 percentage in the alveolar air becomes greatly reduced so as to maintain the tension at around the normal value of 40 mm Hg. Similarly, at low barometric pressures, the alveolar CO_2 tension remains unchanged unless the atmospheric pressure falls to a value of 550 mm Hg or less (equivalent to an altitude of about 10,000 feet) and an O_2 pressure of approximately 60 mm Hg when a sharp fall to around 30 mm Hg in alveolar $p\text{CO}_2$ occurs.

The stimulation of respiration by the inhalation of carbon dioxide ceases when the concentration is about 9 per cent. From then on its *narcotic* or *anesthetic* action becomes evident and the pulmonary ventilation is reduced, with higher percentages the depression of the respiratory mechanism becomes more and more pronounced and cessation of breathing occurs at between a concentration of between 35 and 40 per cent. All narcotics and anesthetics are depressing to the respiratory center as demonstrated by its response to CO_2 inhalation, though it remains sensitive for a time to impulses from the peripheral chemoreceptors which are much more resistant. Dripps and Drumke found that whereas all the narcotics and anesthetics investigated—ether, cyclopropane, chloralose, morphine, barbitol and pentobarbital—depressed the respiratory center, only the first two decreased the sensitivity of the peripheral chemoreceptors.

Both central and peripheral chemoreceptors after a time become tolerant to a moderately high $p\text{CO}_2$. In pulmonary emphysema, for example, the carbon dioxide content of the blood is elevated, and the patient does not respond in the usual way to the inhalation of air containing up to 7 per cent of carbon dioxide. In observations upon the crew of a submarine ("snorkel" type) which did not actually surface for four weeks, the breathing, though stimulated at first, returned to normal though the carbon dioxide of the inspired air was over 5 per cent.

Effect of changes in pH Since changes in CO_2 tension tend to alter the blood pH it is difficult to separate the effects of $p\text{CO}_2$ and of acidity itself upon respiration. But the maximum effect upon pulmonary ventilation in man caused by the inhalation of CO_2 is about 70 liters, whereas that resulting from a fall in pH due to fixed acids, as in certain pathological conditions, is only 35 liters (see table 33 and Nielson Experiment, p. 409). The sensitivity of the central chemoreceptors to

TABLE 33
(After Gray)

CONDITION	MAXIMAL VENTILATION	CHANGES IN ARTERIAL BLOOD			PC*	CC
		$p\text{O}_2$	$p\text{CO}_2$	pH		
	<i>l/min</i>					
Anoxia	12	↓	↓	↑	+	—
Inhalation of CO_2	70	↑	↑	↓	+	+++
Acidosis (metabolic)	35	↑	↓	↓	+	+
Moderate exercise	50	0	0	0	—	—
Severe exercise	120	0	↓	↓	—	—

* PC and CC indicate peripheral and central chemoreceptors, respectively

the H-ion as compared with the sensitivity of the carotid and aortic bodies is a controversial question, but it appears that the center is the more responsive.

The hyperpnea of muscular exercise It will be seen from table 33 that of all the conditions listed, the greatest increase in pulmonary ventilation (a maximum of 120 liters) occurs in severe muscular exercise. But Barcroft and Margeria found that in exercise a less intense sense of effort was experienced by the subject than when the hyperpnea, though not so great, was induced by the inhalation of CO_2 . It is apparent from the table that none or all three chemical factors can be the cause of the powerful effect of exercise upon respiration. The pCO_2 of the blood is not increased in moderate exercise, is actually reduced in severe exercise, and neither moderate nor severe exertion is associated in health with a fall in the pO_2 of the arterial blood. The arterial pH shows a fall (0.1 to 0.4 pH units) only in the severest exercise due to the entrance into the blood of lactic acid (with the formation of lactates) from the active muscles, in moderate exercise no fall in pH, or only a temporary one, can be detected. Furthermore, in strenuous exertion neither the concentration of blood lactate nor the fall in pH shows any definite correlation with the increase in lung ventilation. After exercise, though the pulmonary ventilation has been reduced by half, the arterial pH may be lower than during the exertion. DeMar, a famous runner, had in one instance respired 90 liters per minute with only a slight increase in blood lactate.

Reflex mechanisms, especially those initiated from the active muscles and moving joints, and in the great veins and lungs, appear to be of greater importance than chemical factors in stimulating respiration in muscular exercise. Passive movements in the lower limbs of dogs, severed at the hip except for the intact sciatic nerve and femoral vessels, caused an increase of from 6 to 125 per cent in pulmonary ventilation (Harrison, Comroe and Schmidt). The manner in which such reflexes are initiated, whether by the stimulation of chemoreceptors in the muscles by metabolic products or by a humor carried to the respiratory center, is unknown. Section of the cord and occlusion of the vessels coming from the contracting muscles have given conflicting evidence.

Other factors to be thought of in the search for the cause of the extreme hyperpnea of exercise are the liberation of adrenaline the rise in tem-

perature of the blood which may stimulate respiration through central thermoreceptors and hypothalamic-pneumotaxic connections, as well as reflexly through cutaneous thermoreceptors. The hyperpnea occurs so promptly after the beginning of the exercise that Krogh has suggested that the respiratory center is sensitized by impulses irradiated from the motor area of the cortex. The rise in venous pressure in the great veins and the stimulation of receptors in the lungs are also no doubt important factors. But it must be said that no known predominant factor can explain the hyperpnea of exercise for none alone causes a sufficiently intense effect. Even all known effects combined cannot account for the enormous increase in pulmonary ventilation caused by exercise. Grodins, after an analysis of all factors, chemical and reflex, postulates an undiscovered *exercise stimulus* which in its intensity is directly proportional to the metabolic rate.

A summary of present day views on respiratory regulation In the light of the present state of knowledge it appears that carbon dioxide, as originally proposed by Haldane, acting in some unique and specific manner upon the chemoreceptors of the respiratory center is the chief and essential factor in ordinary quiet (eupneic) breathing with the body at rest or engaged in light exertion. That carbon dioxide acts specifically and not simply because of its acidic properties is supported by the experiment of Nielson.

In other than eupneic respiration, e.g., anoxia, exposure to high temperature, acidosis due to fixed acids, etc., other factors—chemical and reflex—are called into play. But no simple theory, stressing one or other chemical factor (pO_2 , pCO_2 , or H^+) or reflex mechanism is satisfactory. Thus, in anoxia there is a fall in pCO_2 and a rise in pH, in acidosis there is a fall in pCO_2 and a rise in pO_2 , and in the hyperpnea of CO_2 inhalations, though two of the three factors change in the right direction there is a rise in pO_2 . Recognizing and emphasizing these facts, Gray has formulated what he terms a multiple factor theory, which postulates the action in its own right of each of the three chemical factors, pCO_2 , H^+ and pO_2 , yet their concentrations in the blood are interdependent in that any change in the concentration of one is accompanied by a change in that of the others, either in the same or opposite direction. That is, different effects of the three chemical factors may be additive or antagonistic, the final result being the algebraic sum of their

partial actions. Thus, the hyperpnea of acidosis causes increased elimination of CO_2 and a consequent fall in pCO_2 of the blood, with the result that the acidosis is less pronounced than it would be had no increase in ventilation occurred, in anoxia the greater pulmonary ventilation induces a CO_2 deficit and, as a consequence, a rise in pH, both effects tending to moderate the result upon the respiratory mechanism of the fall in pO_2 .

HYPERPNEA, FORCED BREATHING AND APNEA

Hyperpnea An increase in the quantity of air breathed (minute volume) as a result of an increase of either the rate or depth of respiration or of both, is called hyperpnea. Hyperpnea may be produced by impulses reaching the respiratory center from the cerebral cortex (as in excitement and other emotional states) or from the hypothalamus, by the stimulation of sensory nerves (e.g., pain, heat or cold applied to the skin, etc.), by conditions which increase the demand of the tissues for oxygen, e.g., muscular exercise, and by certain other factors to be mentioned in the section under dyspnea.

Apnea If the lungs of an animal are over-ventilated for a minute or two, either by stimulating a sensory nerve and thus inducing reflex hyperpnea or by means of some mechanical respiratory device, a period follows during which all breathing is suspended. This period of respiratory arrest is called *apnea*. An apneic period can also be induced by the stimulation of the pressoreceptors in the carotid sinus as by the sharp rise in blood pressure produced by an injection of adrenaline.

Any person can readily induce apnea in himself by overbreathing. When one breathes deeply and quickly for a few minutes and then stops he does not have any desire to breathe for the next minute or more. The deep and rapid breathing is termed voluntary hyperpnea or forced respiration.

The apnea following overventilation is due to the excessive elimination of carbon dioxide and in consequence to the lowering of the tension of this gas in the blood. A fall in alveolar carbon dioxide tension to 15 mm Hg can be demonstrated in man after a short period of over ventilation. The alveolar pCO_2 rises gradually during the apneic period and reaches a value a little below normal before breathing is resumed. Forced breathing in an atmosphere rich in carbon dioxide therefore does not cause apnea. Owing to the shape of the oxygen

dissociation curve of hemoglobin, oxygenation of the arterial blood can be only slightly improved by forced breathing, and is therefore of little significance in the production of the apnea. Indeed, in the apneic periods of Cheyne Stokes respiration there may actually be oxygen want sufficient in degree to cause pronounced cyanosis (Haldane and Poulton), yet the apneic period continues until a further rise in pCO_2 occurs. On the other hand when forced breathing is carried out in an atmosphere containing a high percentage of oxygen apnea may persist for from 6 to 8 minutes, or even more, and, in one instance reported by Schneider, for 15 minutes! The high percentage of oxygen in the inspired air causes a slight rise in the oxygen tension of the body fluids and an increase in the total oxygen content of the body. This extra oxygen store is drawn upon during the apneic period. The resumption of breathing following the apnea of forced breathing of ordinary air is thought therefore to be due in part to anoxia (stimulation of the carotid and aortic bodies).

Other effects of forced breathing

While the cardinal features of the reaction to forced breathing are now well known, there is considerable variety in detail when observed over a large number of cases. The first sensations of the subject are usually dizziness and faintness resulting in some relaxation of his efforts, but it is very rarely that a feeling of nausea becomes prominent.⁵ This initial discomfort may arise from a disturbance of the circulation directly due to the too vigorous movements of the chest. Within five minutes at the most, a regular rhythm has been established and the subject will have begun to notice numbness and tingling, probably coolness and perhaps tremors in various parts of the body and slight tightening of the muscles. In the majority of cases, samples of alveolar air taken at this time will show a level of carbon dioxide about one half the normal, and only a slight further drop will occur as the experiment continues. The pulse and blood pressure will have risen, as would be expected with muscular exercise, but from now on signs and symptoms will begin to vary. The pulse and blood pressure may retain their relationship, or the pulse may rise without corresponding rise in pressure, but not infrequently the pulse rate increases while the pressure falls. It is usually in this last group of cases that one may find the cold clammy skin, pale cyanosis and thready pulse which form a curious anomaly. If enough hemoglobin is reduced to cause cyanosis it obviously is not holding its oxygen by reason of the alkalosis,

⁵ These observations on voluntary hyperpnea were kindly placed at our disposal by our colleague, Dr Edward Fidler.

and therefore the circulation must in some manner be restricted. The loss of carbon dioxide may interfere with the efficiency of the heart. There is evidence of vasoconstriction of the peripheral vessels, and vasodilation of the vessels in the splanchnic area has been observed during surgical operations. Good evidence of circulatory disturbance exists in the coldness of the extremities. The condition is suggestive of shock.

Mild tetany occurs in a certain proportion of subjects due to a shift in the acid base balance toward the alkaline side. The tetanic manifestations vary from a slight stiffness of the muscles to well-marked contractions. In the hand the most common attitude is ulnar flexion with relative extension on the radial side and some flexion of the wrists. Before it is clearly evident it may occasionally be brought out by constricting the upper arm (Trousseau's sign).

The hyperexcitability of the nervous system is almost always shown in the increased briskness of the knee jerk, but particularly in the facial nerve reflex elicited by tapping just below the zygoma (Chvostek's sign, ch 60).

It is after the first five minutes that certain temperamental differences may also appear. Whereas many, perhaps more cautious, keep themselves well in hand, cooperate easily in the taking of the samples and can stop the moment the signal is given, there are others, perhaps more adventurous, who develop symptoms more quickly, have a little difficulty in giving samples and finally have trouble in coming to a halt. Of the latter group some become somnolent and a few will appear amused and if set laughing have difficulty in controlling themselves. The condition has been described as one of mild intoxication, and in class-room experiments immediately after the war a student who had been an aviator remarked that his sensations resembled those resulting from a very rapid rise in altitude (see p 422).

The apneic period is variable in length depending on the degree of disturbance and is, of course, prolonged if oxygen is taken during the last few breaths of the forced breathing. Whether cyanosis has appeared before or not it is almost certain to be present with an apnea of more than thirty seconds. During long

periods the change may be very striking where the texture and color of the skin enhance the contrast.

The return of breathing occurs more frequently with a single breath, usually rather small, followed by a secondary apnea. Succeeding movements gain in extent while apneic periods decline until normal breathing is reached. Typical Cheyne-Stokes breathing, which at one time was considered a common characteristic, has been quite rare.

PERIODIC RESPIRATION

Periodic breathing is the term applied to various types of uneven respiratory rhythm. The most common type is called Cheyne-Stokes breathing after two physicians who described it (It had been described earlier, however, by John Hunter, the famous London surgeon of the 18th century). In this type of breathing the respirations may be described as alternately waxing and waning. That is, a period of breathing occurs in which the individual respirations are small and slow to start with but gradually increase in depth and rate to a maximum and then, subsiding again, finally cease for a time (fig 33 9). The apneic phases last for about 35 seconds. Periodic breathing of this type occurs at high altitudes and is exhibited by such animals as the ground hog and the dormouse during hibernation. A tendency to this type of breathing is not uncommonly shown by healthy infants and occasionally by normal adults during sleep. The most common clinical conditions in which Cheyne-Stokes breathing occurs are advanced renal and cardiac disease, asthma and raised intracranial pressure. It is also seen in severe pneumonia and in morphine and chloral poisoning, or it may follow a general anesthetic. As mentioned above, it occurs, though rarely, after a period of forced breathing. Douglas and Haldane induced it in normal persons by a moderate degree of oxygen want. Clinical Cheyne-Stokes respiration is usually a grave omen and is probably initiated, in most

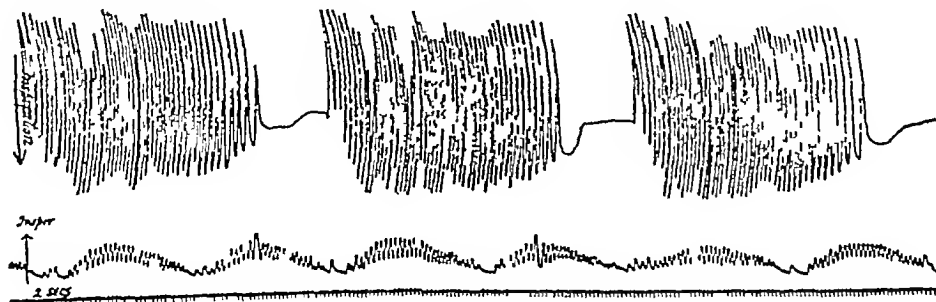


FIG 33 9 Two examples of Cheyne-Stokes breathing (After Lewis)

instances at any rate, by damage to the respiratory center caused by oxygen lack, it is abolished in many cases by the inhalation of oxygen. It apparently develops in the following way. The depressed state of the respiratory center results in weak, shallow respirations which intensify the existing anoxia. The oxygen lack when it reaches a certain degree, together with the accumulation of carbon dioxide in the blood, stimulates the respirations, oxygenation of the blood is improved, and the $p\text{CO}_2$ of the arterial blood is reduced below the threshold for stimulation of the center either directly, or through the chemoreflex mechanism, the apneic period ensues, and, of course, again increases the anoxia and prevents the elimination of CO_2 , the breathing returns and the cycle is repeated. It has also been pointed out by Douglas and Haldane that the elimination of large quantities of CO_2 from the body during the hyperpneic periods and the ultimate reduction of the bicarbonate reserve removes a steadying influence which is normally exerted upon the respiratory center. (They compare this effect to that of the fly-wheel of a motor.) The body normally holds a large store of CO_2 in the lungs and tissue fluids (as bicarbonate) which is drawn upon to oppose any sudden fall in CO_2 tension of the body supplying the respiratory center. Any sharp increase in CO_2 tension, on the other hand, is prevented through the excess CO_2 being buffered by the tissue fluids. In Cheyne-Stokes breathing the reduction of the bicarbonate reserves permits slight changes in the gas tensions to produce sudden and exaggerated fluctuations in the activity of the center. CO_2 administration as shown by Allen and Pembrey abolishes in many instances the periodic rhythm.

A manner in which periodic breathing might possibly be produced in cases of raised intracranial pressure is suggested by an experiment of Eyster. He increased the intracranial pressure in dogs and found that when the arterial blood pressure rose and fell alternately above and below the intracranial pressure, Cheyne-Stokes breathing ensued. In clinical cases of raised intracranial pressure similar alternating elevations and depressions of the blood pressure have been observed by Eyster, the hyperpneic phase coinciding with the blood pressure rise. The blood pressure changes may be the immediate cause of the respiratory rhythm in these cases. It is easy to conceive that a fall in arterial pressure below intracranial pressure by causing acute anemia of the center would suppress its activity for the time and induce apnea, this would give place to hyperpnea during the phase of raised arterial pressure when blood again filled the vessels of the center.

A type of respiratory periodicity known as Biot's breathing is seen in disease, e.g., meningitis, involving the region of the medulla. This differs from the Cheyne-Stokes type in that the onsets of the apneic and hyperpneic phases are abrupt. The two phases are frequently also unequally spaced, the breathing being less regularly rhythmical.

DYSPNEA

Definition

Dyspnea literally means difficult breathing. When the respirations from whatever cause cannot be carried out with ease and practically unconsciously the individual is said to be dyspneic. The term therefore implies a subjective element. Dyspnea thus differs from hyperpnea. The latter term means simply increased pulmonary ventilation, and this may occur quite unconsciously, or if the subject is aware of the augmentation of the breathing there is not necessarily any sensation of difficulty or distress. When the hyperpnea becomes extreme and yet leaves the need for which it had been instituted unsatisfied, discomfort or distress is experienced and the term dyspnea is applicable. Meakins offers the following concise definition: "Dyspnea is the consciousness of the necessity for increased respiratory effort".

THE CAUSES OF DYSPNEA

Since the respiratory and circulatory functions are directed toward the acquisition of oxygen and the elimination of carbon dioxide, dyspnea may result if either of these functions be disturbed to such an extent that the normal gaseous exchanges cannot be accomplished. On the other hand the oxygen requirement and the carbon dioxide production may be so great that the *normal* respiratory and circulatory mechanisms find difficulty in meeting the demands of the moment. Or again the supply of oxygen itself may be inadequate as a result of a low oxygen tension in the atmosphere (e.g., high altitudes). In considering the causes of dyspnea we must therefore in many instances look beyond the lungs themselves. The lungs are the bellows but it is the tissues which consume oxygen and produce carbon dioxide, and it is the heart and the blood which are concerned with the carriage of these gases between lungs and tissues. As already mentioned dyspnea is the sensation of respiratory distress. The height of the hyperpnea at which dyspnea appears is called the *dyspneic point*. There is a relationship between the latter and the *vital capacity*. A person with a large

vital capacity obviously can breathe a larger volume of air without discomfort than can one with a smaller vital capacity. There is a close relationship however between the onset of dyspnea and the ratio of the functional residual air (that is, the quantity of air in the lungs at the end of an ordinary expiration) to the total lung capacity, (i.e. the total volume of air which the lungs can hold during full inspiration (p. 360)). In health the ratio is about 2 to 5. The more nearly equal are the volumes of the functional residual air and the total lung capacity, the greater will be the tendency to dyspnea.

The fundamental or immediate causes of dyspnea can in most instances be reduced to the following categories: (1) Stimulation of the respiratory center either (a) reflexly from the carotid and aortic bodies (anoxia or increase in H-ion concentration due to fixed acids), (b) directly, by CO₂ excess, (c) a combination of (a) and (b) as in asphyxial states, or (d) by impulses from cerebral centers or by afferent impulses, especially of a painful character, from abdominal or peripheral regions. (2) Hypersensitivity of the Hering-Breuer reflex, thus bringing about earlier inhibition of the inspiratory phase and causing as a consequence a more rapid shallow type of breathing.

The abnormal conditions chiefly associated with dyspnea are

(1) Prevention of adequate oxygenation of the blood in the lungs. That is, anoxia of the anoxic type (p. 419). This may result from (a) pulmonary disease or from laryngeal or tracheal obstruction, or (b) low O₂ tension in the inspired air, e.g., high altitudes.

(2) Interference with the transport of the respiratory gases. Anoxia of the stagnant or anemic type (a) slowing of the circulation as in cardiac failure, (b) severe anemia, breathlessness chiefly on exertion.

(3) Restriction of the action of the diaphragm or intercostals.

(4) Acidosis, reduced alkali reserve or retention of CO₂ (gaseous acidosis)

(5) Increased metabolism ✓

(6) Nervous conditions, e.g., emotional disturbance, neurasthenia, hysteria, encephalitis, or the direct stimulation of the respiratory center by cerebral tumor, hemorrhage or edema.

Dyspnea due to pulmonary diseases

Dyspnea is a feature of various respiratory diseases

(a) In some instances, e.g., laryngeal or bronchial obstruction and asthma, the dyspnea is due to a combination of anoxia and CO₂ retention.

(b) In other instances owing to the reduced distensibility of the lungs resulting from edema, congestion, inflammation, fibrosis, etc., the Hering-Breuer reflex is abnormally sensitive.

(c) Limitation of the movements of the diaphragm and chest wall. In emphysema, for example, owing to the loss of lung elasticity the resting position of the chest is one of nearly full inspiration. The diaphragm is fixed and the thorax elevated. Any further enlargement of the chest entails unusual effort on the part of the intercostal muscles (see also p. 351) and the enlistment of the accessory muscles of respiration. Expiration involves active contraction of the expiratory muscles.

Cardiac dyspnea

Dyspnea upon exertion is a feature of certain chronic pulmonary and heart lesions, e.g., mitral stenosis. Stimulation of the carotid and aortic bodies by oxygen want or of the respiratory center by carbon dioxide excess is not, in the absence of cardiac failure, responsible for the dyspnea, since the oxygen saturation of the arterial blood may not be reduced to any important degree and the carbon dioxide tension is within or even below the normal range. Pulmonary engorgement leading to diminished distensibility of the lung is considered by Meakins, Christie and associates to be the prime cause of cardiac dyspnea. Though the reduction in the vital capacity may be roughly proportional to the dyspnea, the two do not bear the relationship of cause and effect, since the subject's vital capacity is always greater than the volume of air required for the exertion which causes the dyspnea. Owing, however, to the diminished distensibility—the stiffness of the lung—a greater inspiratory effort is expended in breathing the extra volume of air which the muscular exertion demands. The lung might be compared to stiffened bellows leather, more force is required to distend it. The elasticity of the lung is also moderately reduced so that expiration instead of being a passive act brought about largely by the recoil of the lung now requires the aid of the contraction of expiratory muscles in order to “squeeze” the air from the chest. The intrapleural pressure, therefore, instead of remaining “negative” throughout the respiratory cycle becomes positive toward the

end of expiration (Christie and Meakins) The decreased distensibility of the lung, for the same reason that it increases the difficulty of enlarging the volume of tidal air, reduces the vital capacity. In other words the dyspnea and reduced vital capacity are due to a common cause. The reduced distensibility will also have the effect as already mentioned, of increasing the sensitivity of the Hering-Breuer reflex with the production of shallow breathing.*

In congestive heart failure with marked slowing of the circulation there is commonly hyperpnea and dyspnea even during rest and then there may be added to the pulmonary factor itself the stimulating effect of carbon dioxide excess upon the respiratory center but, according to Christie and Meakins, this is of minor importance. When pulmonary edema supervenes, interference with the absorption of oxygen and the production of arterial anoxemia (stimulation of carotid body) may possibly be a factor, though we have seen that the peripheral chemoreceptors are not very sensitive to anoxia and the central ones are actually depressed. Hindrance to the absorption of oxygen, caused by the presence of exudate in the alveoli and the edematous swelling of the alveolar walls, is accompanied by little or no interference with the elimination of carbon dioxide, this is probably due in part to the much greater solubility of carbon dioxide than of oxygen and, in consequence, to the freer diffusion of the former gas through the edema fluid. In congestive heart failure arterial anoxemia with a normal or even a subnormal carbon dioxide content of the arterial blood may exist even in the absence of pulmonary edema.

Experimental support can be cited for the view that reduced distensibility of the lungs as a result of congestion is an important factor in cardiac dyspnea. It has been mentioned elsewhere (p. 426) that distension of the pulmonary bed causes rapid shallow breathing. The production of multiple emboli in the pulmonary capillaries by the intravenous injection of starch granules causes congestion of the lungs and rapid shallow breathing (p. 427) and Partridge found that after rapid breathing has been induced in this way the impulses recorded from the vagus nerve are of higher frequency than those resulting from inflation to an equal degree of normal lung. It has also been

shown that in man pulmonary congestion does actually reduce the distensibility of the lung tissue.

Harrison and his colleagues found in subjects of cardiac failure the CO₂ content of the jugular blood to be within normal limits and observed no significant reduction in the cerebral blood flow (i.e., through the respiratory center). They suggest that afferent impulses initiated from the great veins at the base of the heart, as a result of the high venous pressure, as well as impulses from the congested lungs, excite the respiratory center.

Though there is much to be said for the reflex origin of cardiac dyspnea, not all are agreed as to its paramount importance. McMichael, for example, in a clinical study found a reduction in cardiac output in those subjects showing hyperpnea (which is always associated with the dyspnea) during rest. The hyperpnea showed a closer correlation with the cardiac output than with the vital capacity which, he points out, is contrary to what might be expected were congestion of the lungs the dominant causative factor. He is inclined to believe that the hyperpnea and dyspnea of the cardiac patient during rest is due to reduced blood flow through the center, resulting in a high CO₂ tension, and possibly to the accumulation of acid products of its own metabolism.

Other minor factors which may contribute to the dyspnea are the increased metabolic rate in congestive heart failure due in part to the greater respiratory effort, and the defective heat dissipation resulting from the slowed peripheral circulation, some degree of fever being not unusual.

ORTHOPNEA. In congestive heart failure with dyspnea at rest the breathlessness is usually more pronounced in the recumbent than in the sitting position. When propped up with pillows the patient may be quite comfortable but becomes dyspneic when he lies down. Many theories have been advanced in attempts to explain the less difficult breathing in the upright position. Among these are

(a) Removal of the weight of the abdominal viscera which interferes with the descent of the diaphragm in the recumbent position.

(b) Reduction in the intracranial venous pressure and the improved draining of blood from the medulla and in consequence, augmentation of the flow through the respiratory center. Mere bending of the head forward (when recumbent) which reduces the cerebral venous pressure, but not the pulmonary engorgement, reduces the volume of

*The beneficial effects of morphine upon cardiac dyspnea are very probably brought about through the reduction in the sensitivity of the vagal endings.

respired air in the orthopneic patient (Battro and Labourt)

(c) Draining of blood from the chest and the relief of pulmonary congestion. This is probably the most important factor. The distensibility of the lung is thereby increased and the Hering-Breuer reflex rendered less sensitive. The vital capacity is less in the recumbent than in the sitting posture. This is true even for the normal person but in cardiac cases the effect is accentuated by the vascular engorgement and decreased distensibility of the lung induced by recumbency.

CARDIAC ASTHMA This is the term applied to paroxysmal attacks of dyspnea occurring, usually at night, in subjects of heart disease associated with hypertension and advanced arteriosclerosis. The upright position tends to relieve the dyspnea. The cause of the attack is unknown, though Christie and Meakins suggest that it is due to pulmonary congestion and a sudden decrease in pulmonary distensibility. During the attack, according to Weiss and Robb, there is engorgement of the pulmonary vascular bed, a tendency to pulmonary edema and a marked reduction in the oxygen saturation of the arterial blood.

Dyspnea in anemia

When at rest, the anemic subject is as a rule not dyspneic. The hemoglobin though reduced in amount becomes fully saturated with oxygen in the lungs. The oxygen tension and consequently the quantity of the gas in simple solution in the arterial blood are normal. The arterial blood of a patient whose hemoglobin is 30 per cent of the normal value will, however, contain only a little over 6 volumes per cent of oxygen. In the healthy resting body the blood in its passage through the capillaries gives up about 5 volumes per cent. If the velocity of blood flow through the tissues in anemia were the same as during health a unit volume of blood would give up an equivalent amount of oxygen. This would leave a reserve of only 1 volume per cent, i.e., the venous blood would be almost completely reduced. The tissues, including the carotid and aortic bodies, would suffer at all times from anoxia. The circulation rate (cardiac output) is, however, increased in anemia so that each unit of blood gives up a smaller proportion than this of its oxygen load. For this reason the chemoreceptors are not stimulated and the patient is not dyspneic while resting even

though his hemoglobin is greatly reduced.⁷ Not only is the output of the heart increased but a redistribution of the blood occurs. The vessels of the skin are constricted and a greater proportion of the total blood volume is driven through other regions. The extent to which the circulatory readjustments can compensate for the blood defect is limited, therefore during exertion the extra demand for oxygen cannot be met. Oxygen want follows, the respiratory mechanism (carotid body) is stimulated, hyperpnea and dyspnea result.

It is to be remembered that in anemia the carriage of carbon dioxide may also be interfered with since hemoglobin constitutes an essential part of the mechanism provided for the transport of this gas (see chapter 32). A deficiency of carbonic anhydrase in the blood, which is carried only by the red cells, and the consequent accumulation of carbon dioxide during exertion has been suggested as an additional cause of dyspnea in anemia.

Dyspnea due to increased metabolism

IN HEALTH, muscular exercise is the outstanding cause of a great increase of metabolism. Carbon dioxide is produced in excessive amounts both from oxidative processes and by the interaction of lactic acid with bicarbonate and acts as a powerful stimulus to the respiratory center. The increased pulmonary ventilation in exercise is due in part only to stimulation of the center by CO_2 .^① Reflexes initiated in the active muscles, and from the walls of the great veins as a result of the high pressure of blood entering the right side of the heart, and possibly as well the irradiation of impulses from the motor cortex, also play an important part (p. 411). As the severity of the exercise is increased, hyperpnea merges into dyspnea. The athlete and the untrained person differ widely in respect to the degree of muscular exertion which will produce this physiological type of dyspnea. The difference depends upon the following factors.

(a) *Vital capacity* In the average normal man the pulmonary ventilation increases from 4 to 5 fold before the dyspneic point is reached. The athlete on the other hand, since his vital capacity is greater shows a correspondingly greater increase in his pulmonary ventilation before dyspnea supervenes. The existence of any pulmonary condition which reduces the vital capacity will depress the level of the dyspneic point.

⁷ Fahr and Ronzone reported a case in which the hemoglobin was 12 per cent of the normal and the arterial blood contained only 2.2 volumes per cent of oxygen. There was no dyspnea during rest.

(b) *Circulation rate* The trained man can increase his circulation rate to a greater degree than the untrained and so deliver more oxygen to his tissues at high pressure.

(c) *Neuromuscular integration* Co-ordination of the several muscles in a given muscular act is more precise in the trained than in the untrained man. There is thus less waste of effort. In the performance of a given amount of work, therefore, the untrained man consumes a greater volume of oxygen, i.e., he is a less efficient machine. Yet his respiratory center appears to be more sensitive to nervous influences—impulses from the contracting muscles and from the cerebral cortex.

A PATHOLOGICAL INCREASE IN METABOLISM, e.g., hyperthyroidism, fever, etc., cannot apparently in the absence of some respiratory or circulatory abnormality cause dyspnea during rest. The increase in metabolism would need to be around 300 per cent before dyspnea might be expected to occur and such a metabolic level is never reached in these or any other pathological condition. Nevertheless, in hyperthyroidism dyspnea will follow upon a degree of exertion which would cause no distress in a healthy person, for the greater metabolic rate, due to the disease, added to that of the exercise will increase the metabolism sufficiently to raise the pulmonary ventilation above the dyspneic point. The dyspnea of hyperthyroidism, when pulmonary and cardiovascular complications are absent is therefore like that of anemia, evident only upon exertion. In hyperthyroidism there is also a reduced vital capacity (which will lower the dyspneic point) as well as a diminution of muscular efficiency.

Dyspnea due to acidosis

The importance of the part played by pulmonary ventilation in resisting a rise in the hydrogen ion concentration of the body fluids has been dealt with elsewhere (ch. 13). Little more need be said here. Non-volatile acids, e.g., lactic in muscular exercise, β -hydroxybutyric and acetoacetic acids in diabetes, and retained acids in

nephritis, react with bicarbonate. The alkali reserve becomes reduced, but the ratio,

$$\frac{\text{H}_2\text{CO}_3}{\text{NaHCO}_3} = \frac{1}{20}$$

is maintained through the stimulating effect of CO_2 upon the respiratory center. It is thus that the CO_2 of the alveolar air (and so of the arterial blood) is kept at a level proportional to the reduction in the denominator of the equation. When this can no longer be effected, i.e., when the hydrogen ion concentration of the blood rises, the acidosis being then uncompensated, the center is stimulated both directly and through reflexes initiated from chemoreceptors of the carotid and aortic bodies.

J. B. S. Haldane produced a severe acidosis and dyspnea in himself by the ingestion of acid forming salts (CaCl_2 and NH_4Cl). But as a matter of fact, the production of fixed acids in diseased conditions rarely causes such a high degree of hyperpnea that dyspnea results, unless the circulatory and respiratory mechanisms are very inefficient or the metabolic rate is increased. According to Means the "alkali reserve" must drop to around 12 volumes per cent before dyspnea supervenes as a result of the acidosis itself. In milder grades of acidosis, however, dyspnea, as in hyperthyroidism occurs upon exertion. The hyperpnea due to the increased metabolism of the exercise is then added to that due to the acidosis with the result that the dyspneic point is soon reached.

Carbon dioxide retention may, as in emphysema, be accompanied by a compensatory rise in bicarbonate and a normal blood reaction (compensated CO_2 excess), dyspnea is not a notable feature in this condition. In other instances of CO_2 retention, compensation is incomplete (uncompensated CO_2 excess or gaseous acidosis) and dyspnea results. In others, again, the CO_2 retention is due to the depression of the respiratory center itself, as in morphine narcosis, in such instances, though compensation is incomplete, dyspnea of course does not occur.

CHAPTER 34

ANOXIA (MORE CORRECTLY HYPOXIA)

CLASSIFICATION AND DEFINITIONS

The failure of the tissues, for any reason, to receive an adequate supply of oxygen is called *anoxia*, *hypoxia*, *oxygen lack* or *oxygen want*. Anoxia may result from (a) defective oxygenation of the blood in the lungs, (b) lowered oxygen capacity of the blood, or (c) slowing of the movement of blood through the capillaries. Upon the basis of these three causes, oxygen want has been classified by Barcroft into *anoxic*, *anemic* and *stagnant* types, respectively. A fourth type, the *histotoxic* (Gk *histos*, tissue, and *toxikon*, poison), caused by poisoning of the oxidative processes of the tissues has been added by Peters and Van Slyke. *Anoxemia* is a term which denotes a low tension of oxygen in the arterial blood, and is used synonymously with anoxic anoxia. *Asphyxia* is sometimes used interchangeably with anoxia, but, strictly speaking, this term should be restricted to conditions in which there is anoxia combined with an increased tension of carbon dioxide in the blood and tissues.

THE ANOXIC TYPE OF ANOXIA

Normally, as we have seen, the blood leaves the lungs about 97.5 per cent saturated with oxygen. When it reaches the tissues it contains, therefore, about 19.5 volumes per cent of oxygen when the normal oxygen capacity is 20 volumes per cent. Oxygen is consequently at high pressure in the plasma. During rest only 5 volumes or so per cent of the gas are abstracted, i.e., the mixed venous blood coming to the lungs contains 14.5 volumes per cent (about 73 per cent saturated). As will be seen from the dissociation curve of oxyhemoglobin the supply of this quantity of oxygen is associated with a drop in oxygen pressure to only 40 mm Hg or so. A high pressure gradient from capillary to tissue cells is therefore assured. In the anoxic type of anoxia the saturation of the arterial blood and the oxygen partial pressure are reduced. When the saturation is, say, 50 per cent, there are still 10 volumes per cent of oxygen in the blood available for the tissues. Since the latter require only 5 volumes, 10 volumes apparently should be adequate. But this load is held at a

relatively low pressure, about 25 mm. Hg. After supplying the required 5 volumes per cent, the blood leaving the capillaries will be only about 25 per cent saturated, and the oxygen pressure around 15 mm Hg. The oxygen must obviously be delivered at a very low pressure head. That is, the oxygen pressure gradient from blood to tissue cells is less steep than normally, and the oxygen supply to the tissues, especially to those cells farthest away from a capillary, is reduced. Accompanying the fall in oxygen tension of the blood is a fall in CO_2 tension if there is hyperpnea, and a rise in the pH.

The chief manifestations of the anoxic type of anoxia are an increase in heart rate, dyspnea, cyanosis, mental disturbances, e.g., exhalation, delirium, mania, or fixed ideas. Of all the tissues, those of the central nervous system and the vascular system are the most susceptible to injury by anoxia. Capillary damage leads to a loss of fluid from the circulation. The intensity of the effects upon the body of a low oxygen tension in the arterial blood is influenced by (a) the abruptness of the onset of the anoxemia, (b) its degree, (c) its duration and (d) the general physical condition of the body.

The anoxic type of anoxia, acute or chronic, is produced by the following conditions:

- (1) Low oxygen tension in the inspired air
(a) vitiation of the atmosphere by foreign gases, (b) high altitudes, mountain sickness
- (2) Abnormalities of the pulmonary mechanisms, pneumonia, asthma, emphysema, collapse of the lung, pulmonary engorgement or edema, pneumonia, asthma, water in the lungs of the apparently drowned, obstruction of the air passages, paralysis of the respiratory muscles, or depression of the respiratory center by narcotics and anesthetics
- (3) Direct communication between the right and left sides of the heart through which venous blood is short circuited—shunt

THE ANEMIC TYPE

The anemic type is caused by (a) *hemorrhage* or *anemia* from whatever cause, (b) *carbon monoxide*

poisoning (p 434), (c) poisoning by nitrites and chlorates which like carbon monoxide form stable compounds with hemoglobin (p 57) In the anemic form of anoxia (resulting from anemia (a)) the total load of oxygen is reduced in proportion to the reduction in the hemoglobin The oxygen tension and consequently the amount of the gas in simple solution is, however, the same as in health and the hemoglobin present in the arterial blood is 97.5 per cent saturated In other words what hemoglobin there is, holds its full oxygen load But in order that the tissues shall receive their full quota of oxygen (unless the blood flow through the part increases considerably, for there being less hemoglobin it must load and unload its oxygen more frequently) the blood must give up a larger proportion of its oxygen load than normally (see p 374), and a large part of the oxygen supply will be delivered at low pressure

THE STAGNANT TYPE

In the anoxia due to a slowed peripheral circulation, the saturation of the arterial blood, its total oxygen load and its oxygen tension are all normal A large part of the oxygen supply is, however, delivered under low pressure, since each portion of blood gives up a larger proportion of its load owing to its longer stay in the capillaries Slowing of the circulation from whatever cause will induce the stagnant type of anoxia. It there-

fore occurs in the following conditions (a) heart failure, (b) obstruction of the venous return from a part (local anoxia), and (c) surgical shock Pronounced slowing of the peripheral blood flow can be readily induced in a normal person by the application of cold Meakins and Davies found that the venous blood of an arm vein was completely reduced after the hands had been kept in cold water for a time. Hot water, on the other hand, speeded up the peripheral blood flow, the blood of an arm vein having the oxygen saturation of arterial blood

THE HISTOTOXIC TYPE

In this type, as its name implies, the respiratory mechanisms of the tissues are poisoned The cells are unable to use the oxygen carried to them and, as a result, the capillary and venous bloods contain more oxygen and have a higher oxygen tension than normally Cyanides which inhibit the action of cytochrome oxidase stand first as tissue poisons, but barbiturates, morphine and anesthetics by depressing the dehydrogenase enzyme systems also cause tissue anoxia Edema, too, by interfering with the diffusion of oxygen to the cells, and fever by increasing the oxygen demand, cause, respectively, an absolute and a relative tissue anoxia (see Cyanide poisoning, p 435)

The four types of anoxia are represented in figure 34.1 It will be noted that in the anoxic type

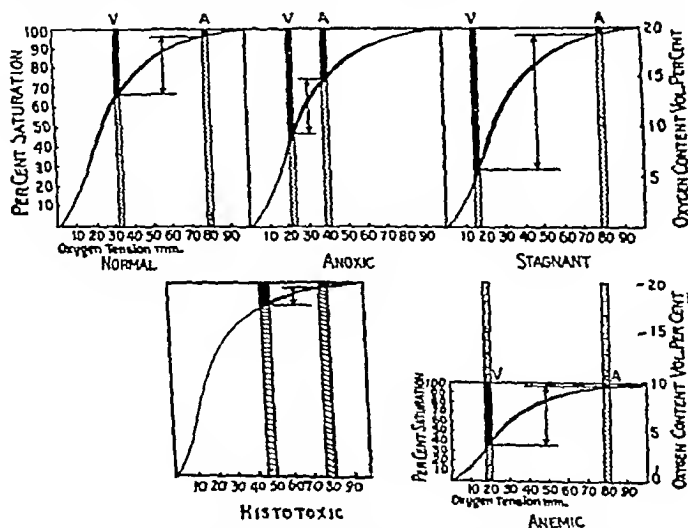


FIG 34.1 Diagram illustrating types of anoxia. Columns representing arterial blood (A) and venous blood (V) are superimposed upon the oxygen dissociation curve. The black portion of the column represents reduced hemoglobin and the shaded portion, oxygenated hemoglobin. In the case of anemic anoxia the dotted portion of the columns represents hemoglobin that is either lost, as in true anemia, or unfit for oxygen transport, as in carbon monoxide poisoning. The perpendicular arrows denote the volume of oxygen given up to the tissues from a unit of blood. (After Means, with additions.)

only, is the arterial saturation low. In the first two types the saturation of the hemoglobin in the venous blood is subnormal. In the anemic type, the hemoglobin, although reduced in amount, is nearly fully saturated, as in health. In the histotoxic type, as a result of the action of the poison upon the oxidative processes of the tissues (ch 32) the oxygen saturation of the venous blood approaches that of arterial blood.

CONDITIONS ASSOCIATED WITH THE ANOXIC TYPE OF ANOXIA

LOW OXYGEN TENSION OF THE INSPIRED AIR

Vitiation of the atmosphere by inert gases

The atmosphere may have a low oxygen tension yet the barometric pressure be no lower than usual. That is, other gases are present which displace the oxygen and so reduce its percentage in the atmosphere. In mines for example the air may contain "Fire Damp" (methane) or "Black Damp" (nitrogen, CO_2) which lower the percentage of oxygen in the air though the barometric pressure is actually higher than that at sea level. These gases, under such circumstances, produce no direct deleterious effect upon the body. They act simply by reducing the oxygen tension of the atmosphere. They therefore differ in this respect from carbon monoxide (p 434), which when occurring in mines in dangerous concentrations is spoken of as "after damp".

Mountain sickness

At high altitudes the percentage of oxygen in the atmosphere is the same as at sea-level but the barometric pressure is low and the partial pressure of oxygen is reduced to a corresponding degree. If the barometer, for example, registers a pressure of 600 mm Hg the oxygen pressure is $(20.96/100 \times 600 =)$ 125.76 mm Hg. Mountain sickness as first shown by Paul Bert is due to the low partial pressure of oxygen in the atmosphere and to the resulting anoxemia, not to the low barometric pressure itself. The effects resulting from breathing atmospheres low in oxygen but possessing a high total pressure are in many respects similar to those of mountain sickness. On the other hand if an animal be placed in a steel chamber and the barometric pressure lowered to $\frac{1}{3}$ or $\frac{1}{4}$ of an atmosphere no ill effects result, provided the air is enriched with oxygen, i.e., if its percentage be raised to maintain its partial pressure at the normal figure.

The first signs and symptoms of oxygen want usually make their appearance in healthy unacclimatized persons when they ascend to an altitude of 10,000 feet or more above sea-level. The height at which mountain sickness appears varies considerably, however, in different individuals depending upon their fitness, the rapidity with which the ascent is made and upon the degree of muscular effort undertaken. From 23,000 to 25,000 feet is around the limit to which unacclimatized men can ascend without the aid of an artificial oxygen supply. Rapid ascents as by plane above 12,000 or 13,000 feet without an artificial supply of oxygen are risky.

Several mountain climbing expeditions have been made in the past by different groups of physiologists for the purpose of studying the effects of low oxygen tensions upon the respiratory functions and of determining the factors underlying the phenomenon of acclimatization. An expedition was made to Monte Rosa (15,000 feet) in 1894 by Mosso and by others subsequently, to the peak of Teneriffe (12,000 feet) in 1910 by Zuntz, Barcroft and associates, and in 1911 to Pike's Peak (14,100 feet) by the Anglo-American expedition of which Haldane, Douglas, Henderson and Schneider were members. In 1921-1922 Barcroft headed a party to Cerro de Pasco (14,200 feet) in the Peruvian Andes. A Himalayan expedition was led by Hartman in 1931, and in 1935 Dill led a party to the Chilean Andes.

It may be useful to the reader if he is reminded of certain fundamental physical principles involved in ascents to high altitudes before considering the physiological effects. As the pressure becomes reduced with increasing altitude the gases of the atmosphere and in the lungs must, of course, expand. It follows that the quantity of oxygen and other gases of the atmosphere per unit volume becomes reduced proportionately. For this reason at great altitudes (above 28,000 feet) the inhalation from a cylinder, by means of tubing and a mask, of even pure oxygen will not prevent anoxia.¹

¹ A rough calculation will make this clear. At 40,000 feet the atmospheric pressure is about 141 mm Hg, i.e., less than $\frac{1}{3}$ the pressure at sea level. A unit volume of air, therefore, contains less than $\frac{1}{3}$ as much oxygen as the same volume at 760 mm Hg pressure. A given volume of pure oxygen would contain less than $\frac{1}{3}$ the number of oxygen molecules at 40,000 feet as it would at sea level. That is to say, a given space at the former level could not contain as much oxygen as a corresponding volume of air would contain at the pressure of 1 atmosphere. Making allowance for the pressures of water vapor (47 mm) and CO_2 (36 mm) the partial pressure of oxygen in the lungs at 40,000 feet while pure oxygen was being breathed would be only $141 - (47 + 36) = 58$ mm Hg.

The pressure of oxygen in the lungs of a person breathing pure oxygen in this way at a height of say 40,000 feet would be only about 60 mm Hg and his arterial blood only 88 per cent saturated with oxygen. A closed cabin containing air at a higher pressure (equivalent to an altitude of around 8,000 feet) is the only way in which the problem of anoxia in high altitude flying can be solved satisfactorily. Even then there is the danger of leaks, especially in military flying from enemy action. Were the airmen thus suddenly exposed to the low pressure existing at altitudes above 35,000 feet, acute anoxia would result, consciousness being lost within a few seconds. The sudden or "explosive" decompression might cause other serious effects (p. 424).

GENERAL SIGNS AND SYMPTOMS. Aeroplane ascents, if made rapidly without the use of oxygen, may result in sudden loss of consciousness due to the reduction in oxygen supply to the brain. When the ascent is made more slowly, or the altitude is not so great as to cause immediate loss of consciousness, the aviator may at first experience sensations of excitement, exhilaration and well-being. As higher altitudes are reached effects of a more serious nature develop, often insidiously. Mental and sensory dullness, muscular weakness, headache, vomiting, cyanosis, dyspnea and perhaps a tendency toward periodic breathing may be induced. A common and dangerous effect is the development of fixed ideas which may result in the performance of foolhardy and ill-judged actions.

When a person climbs to a mountain height the time taken in the journey allows a certain degree of physiological readjustment to take place and the symptoms are usually less intense. But, as in the case of the aviator, mental features, e.g., a feeling of elation, exhilaration, talkativeness and sometimes emotional outbursts, laughing or crying, quarrelsomeness or the development of fixed ideas, are prominent. Mental tasks, e.g., calculations, memory tests, and telling the time from the mirror image of a clock face are performed less efficiently. Similar effects upon the mind are produced upon persons exposed to low oxygen pressures within a steel cabinet. The mental effects as pointed out by Barcroft are not unlike those caused by drunkenness. To quote his words,

"Alcohol affects different persons in different ways so on my journeyings in high altitudes I have seen most of the symptoms of alcoholism reproduced. I have seen men vomit, I have seen them quarrel, I

have seen them become reckless, I have seen them become garrulous, I have seen them become morose. I have seen one of the most disciplined of men fling his arms about on the ledge of a crevasse to the great embarrassment of the guide. I have seen the most loyal companion become ill tempered and abusive to the point at which I feared international complications would arise."—Lessons from High Altitudes

Complete prostration may follow the earlier symptoms. If the individual remains at the high altitude the symptoms pass off after a time, as he becomes acclimatized to the low oxygen tension.

At a simulated altitude of about 20,000 feet (i.e., in a decompression chamber) most subjects experience failing vision, incoordination, and inability to write or to execute simple mental tasks within 15 minutes or so after the extra oxygen supply has been cut off. At 26,000 feet none are able to retain "effective consciousness" for this length of time without oxygen.

CHANGES ASSOCIATED WITH ACCLIMATIZATION. At an altitude of 14,200 feet as at Cerro de Pasco where Barcroft and his party carried out their investigations the barometric pressure is around 440 mm Hg. The partial pressure of oxygen is therefore about 92 mm. The oxygen tension of the alveolar air is not as far below that of the atmosphere as at sea-level, and varied among the greater number of the party from 55 to 60 mm Hg. The closer approximation of the atmospheric and alveolar oxygen tensions at high altitudes is due to the increased breathing which results in a more effective ventilation of the air sacs. The increased respiration is brought about through the action of the lowered oxygen tension in the blood. The oxygen tension is slightly lower in the arterial blood than in the alveolar air which indicates that the passage of the gas is due purely to diffusion and not to an active secretion by the pulmonary epithelium. If this occurred, as has been suggested (Haldane), one would expect the arterial oxygen tension to be higher than that of the alveolar air. Alveolar carbon dioxide tension also, as a result of the increased pulmonary ventilation, is lower than that at sea-level, it varied in different individuals of Barcroft's party from 23 to 29 mm Hg. At 14,200 feet the arterial blood is from 85 to 88 per cent saturated with oxygen (see fig. 32-4, p. 373).

A marked increase in the number of red cells (see p. 11) and a corresponding increase in hemoglobin content of the blood occur at high alti-

tudes The blood volume is also augmented The natives of mountainous regions have a red cell count of from 6 to 8 million per cubic millimeter The greater quantity of hemoglobin of course raises the oxygen capacity of the blood and so tends to counteract its lowered oxygen saturation That is, the total oxygen content of the arterial blood tends in spite of the low saturation to rise Nevertheless, it may not be evident at first sight how a rise in the oxygen capacity is of advantage, for, blood of normal hemoglobin content even when only 80 per cent saturated possesses a quantity of oxygen which is quite adequate for the needs of the tissues It has already been pointed out, however, that the important factor in supplying the tissues is the oxygen pressure gradient between the plasma in the capillaries and the tissue cells So then, if there are a greater number of red cells each will be required to give up less of its oxygen store in passing through the capillaries to furnish a given quantity of oxygen (see anoxia due to anemia, p 419) Consequently the saturation and the oxygen tension of the venous blood will be maintained at a higher level than otherwise would be possible This means that the mean intracapillary oxygen pressure will also be higher, and as a result the tissues are more effectively supplied with oxygen

The reduction in the alveolar carbon dioxide results in a corresponding decrease in the carbon dioxide tension of the arterial blood The ratio $H_2CO_3/NaHCO_3$, which tends to be altered by the loss of carbon dioxide, is adjusted by a decrease in the excretion of acid and ammonia in the urine, a consequent lowering of the "alkali reserve" and depression of the CO_2 dissociation curve (p 396) The actual pH of the plasma changes little if at all Up to about 12,000 feet if any change occurs it is toward the alkaline side, above this level the blood reaction shows little further change or tends to return to normal Lactic acid, which was thought at one time to be produced in excess as a result of the anoxia is actually formed in smaller amounts at high altitudes than at sea-level Even during severe exercise at 15,000 feet and higher altitudes the lactic acid concentration in the blood is lower than during exercise of comparable severity at sea-level

Barcroft and his party observed a shift to the left in the oxygen dissociation curve of hemoglobin, i.e., the affinity of hemoglobin for oxygen was increased The shift in the dissociation curve is ascribed by Barcroft to an increased alkalinity of

the interior of the red cell This increased alkalinity is in turn a direct result of the rise in the number of red cells The buffering power of the blood is increased through the greater facility offered for the action of the "chloride shift" mechanism (p 132) In other words, when a given amount of carbon dioxide is liberated by the tissues it is distributed among a greater number of red cells than under normal circumstances, therefore, the alkalinity of each cell is reduced to a proportionately less extent There have been conflicting reports concerning this question of the shift in the oxygen dissociation curve Some observers have been unable to confirm Barcroft's finding, while others claim that a shift to the right occurs The truth appears to be that up to about 14,000 feet the affinity of hemoglobin for oxygen increases, but at higher levels the dissociation curve tends to assume the form found at sea-level, and at altitudes of 19,000 feet there is a definite shift to the right.

One might suppose that an increased circulation rate would be an important adjustment to the rarefied atmosphere whereby an adequate oxygen supply to the tissues would be maintained, but except for a temporary increase during the first few days no change in cardiac output occurs at altitudes of less than 14,000 or 15,000 feet Above 15,000 feet the greater degree of anoxia results in an increase in the minute volume of the heart. Before acclimatization, the pulse rate during rest increases by from 15 to 20 beats per minute at altitudes between 15,000 and 18,000 feet. At greater altitudes, especially in persons in poor physical condition, the rate may increase above the normal The acceleration of the pulse, according to Barcroft, is a signal of distress flown by the heart laboring under the effects of the anoxia, rather than an indication of an increased minute volume. The blood pressure shows little or no change up to 15,000 feet, a small rise may occur at higher altitudes For the effect of anoxia on the coronary circulation (see p 327)

In the Andean expedition the diffusion coefficient (p 366) of oxygen was not found to be increased to more than a slight extent during acclimatization yet those members of the party who had high diffusion coefficients to start with suffered less from mountain sickness than those in whom the coefficient was low

Those who have lived all their lives at very high altitudes (around 14,000 feet) have a larger vital capacity than dwellers at sea-level Barcroft reports that a native of Cerro de Pasco of 5 feet, 3 inches in height had a chest of a man of 6 feet Moderate altitudes—up to about 7500 feet appear to have little or no effect upon the chest development.

Though anoxia is the most serious effect of high altitudes with which the mountain climber or avia-

tor has to contend, rapid ascents, as by airplane, cause other important physiological disturbances which should be mentioned, namely, (1) expansion of gases in the gastro intestinal tract, (2) aero-embolism (see below) and (3) pressure disturbances in the ears (ch 79)

Expansion of gases in the gastro-intestinal tract

Like the gases of the atmosphere those in the stomach and intestine increase in volume in proportion to the reduction in pressure. Gas having a volume of 1 liter at sea level expands to 2 liters at the pressure (375 mm Hg) existing at 18,000 feet, to 4 liters at 34,000 feet and to 6 liters at 42,000 feet (pressure 128 mm Hg). Distension of stomach and intestine will result unless the abdomen is supported by a belt or by other means, or the gases are freely evacuated. In rapid ascents distress or even severe pain results if there is any hindrance, as by an obstruction in the colon, to the ready passage of flatus (see also p 425)

Caisson disease—Decompression sickness—Aero-embolism—"the bends"—"Explosive" decompression

Caisson disease and the corresponding condition that occurs during rapid ascents to high altitudes may be conveniently described here. Caisson disease occurs in deep sea divers or workers in caissons when they pass too quickly from the high pressure in which they have been confined to the ordinary pressure of the atmosphere.

When an animal is subjected to high atmospheric pressures the amount of the respiratory gases in the blood plasma and certain tissues increases in proportion to the raised partial pressures of these gases in the alveolar air. Very high pressures are well tolerated and no harm is caused by the increased amount of oxygen or nitrogen provided that they remain in solution,

The cause of the symptoms in caisson disease are the bubbles of gas which form when the atmospheric pressure is lowered too rapidly. The bubbles are composed mainly of nitrogen—but also may contain CO_2 and O_2 . The volume of nitrogen in the fluids and tissues of the body is many times that of oxygen, and double that of carbon dioxide. Nitrogen also diffuses less readily than either of the other gases, and is thus less readily eliminated through the lungs. The bubbles act usually as emboli (aeroembolism) and the symptoms produced depend on the site where such emboli lodge—lungs, brain, heart, anterior or posterior spinal roots,

etc.* There may be extensive capillary hemorrhage and, in extreme instances, frothing of the blood may interfere with cardiac action and with the flow of blood in the blood vessels. The quantity of gas absorbed by the blood and tissues during compression is related not only to the pressure to which the body is exposed, but also to the amount of fatty tissue which it contains, for nitrogen is some 5 times more soluble in oil than in water. Carbon dioxide is also somewhat more soluble (50 per cent) in fat than in water. The quantity of these gases absorbed, especially of nitrogen, by a tissue is thus proportional to its fat content. It has been recognized for many years that obese persons are more susceptible to decompression sickness than those who are lean, and a direct relationship has also been demonstrated in animals between the fat content of their bodies and the severity of the effects of decompression. Decompression from a pressure only half as high as that which causes the death of lean guinea pigs will kill fat animals.

In the non fatty tissues, e.g., central nervous system, muscle, lungs, etc., the bubbles form in the blood stream locally or are carried to them from a distance, and the effects are due to the blockage of the small vessels. Only in fatty tissues, e.g., subcutaneous tissue, bone marrow, adrenal cortex, myelin sheath, etc., are they formed extravascularly. They appear within the fat cells, but may then enter the blood vessels and be carried to remote parts. Developing in the myelin sheaths of sensory nerves they cause pain (bends), and temporary paralysis may result from their formation in the sheaths of motor nerves. In the central nervous system, which is relatively poor in fatty material, bubble formation is entirely intravascular.

Symptoms are not produced unless the excess pressure is more than $1\frac{1}{2}$ atmosphere, however rapid is the rate of decompression or however long is the exposure to the higher pressure. Apparently the super saturation of the blood and fatty tissues is not great enough and the volume of nitrogen liberated upon decompression is therefore not sufficient to form bubbles or at any rate to such an extent as to cause damage.

Aeroembolism can be avoided by slow decompression, the excess nitrogen being then gradually eliminated in the expired air. But slow decompression is inconvenient and tedious. The fact just mentioned that rapid decompression from a pressure of a little over 2 atmospheres does not cause symptoms suggested to Boycott, Damant and Haldane that halving the pressure whatever its height, i.e., a reduction from 4 atmospheres to 2 or 6 to 3 could be carried out rapidly and with safety. This is the basis of the method of decompression now generally employed. Decompression is carried out in stages or steps rather than continuously,

* Robert Boyle was the first on record (in 1670) to witness bubble formation in the tissues. He described it in the eye of a snake which he had decompressed by means of his air pump.

the air pressure to which the subject is exposed at each stage being just half the pressure at the preceding stage

In *military airplane flights* the airman may be required to ascend to upwards of 30,000 feet (226 mm Hg) within a few minutes. He is then subjected to rapid decompression, not, as is the deep sea diver, from a high pressure to atmospheric pressure, but from atmospheric pressure to a pressure of about $\frac{1}{3}$ of an atmosphere. But the same general principles apply and similar results follow as those outlined in the preceding section. A reduction in pressure from 1 atmosphere to $\frac{1}{3}$ of an atmosphere would correspond to a reduction from 3 atmospheres to 1 atmosphere. At 20,000 feet (350 mm Hg) the pressure is little more than halved and aeroembolism would not be expected to occur. The effects of aeroembolism in flyers take the form most commonly of severe pain in one or more of the large joints (air-bends), itching of the skin or cutaneous sensations of heat or cold. Other more serious symptoms, e.g., paralysis due to the formation of bubbles in the spinal cord or brain, intense burning pain in the chest, or pulmonary edema may, though rarely, occur. The symptoms are, however, rarely as severe as those occurring in compressed air sickness, the effects are not, as a rule, permanent and fatalities seldom result. They are quickly relieved in most instances by descending at once to lower levels. The reason for the greater severity of decompression effects from high pressures to atmospheric than from atmospheric pressure to high altitude pressure appears to be that, though the actual difference in pressures (i.e., in mm Hg) and the amount of gas capable of release may be the same in both instances, the weight of gas, i.e., the number of molecules per unit volume of gas, released in the former instance is much greater, it will thus exert a greater mechanical effect. Aeroembolism is rare even with very rapid decompression (equivalent to ascents of 12,000 feet per minute) at altitudes below 30,000 feet. But above this apparently critical level they may result from a rate of ascent of only 200 feet per minute. The rate of ascent necessary to induce aeroembolism becomes progressively slower with increasing altitude, at 40,000 feet bends may result from a rate of only 80 feet per minute. The most effective way to prevent or diminish the effects of rapid decompression is to have the airman breathe pure oxygen or an oxygen-helium mixture (oxygen 21 per cent, helium 79 per cent) for several hours before the flight and thus induce the elimination of nitrogen from his tissues.

"Explosive" decompression The profound anoxia which may result if a pilot, as a result of a break in a pressurized cabin, is suddenly exposed to the atmosphere at high altitudes, has been mentioned (p. 422). Other effects which may be produced have been investigated in animals. It has

been found that they have an astonishingly high tolerance to explosive decompression at simulated high altitudes (from sea level to 50,000 feet—750 to 87 mm Hg). The rate of decompression in the experiments of Whitehorn and his associates ranged from about 1100 mm Hg per second to as high as 33,000 mm Hg or more per second. That is to say, decompression from 750 mm Hg to 87 mm Hg was brought about in the latter instance in 0.02 second. The animal was quickly recompressed again and the effects of anoxia thus prevented. The chief effects observed in these experiments were distension of the hollow viscera and lungs, fixation of the expanded thorax in the inspiratory position and of the diaphragm in the expiratory position, due to upward pressure of the distended stomach and intestines, sharp but temporary fall in blood pressure with cardiac slowing, hemorrhages into lungs and hollow organs, or rupture of the latter. The fall in blood pressure, which does not occur for a heart beat or two, seems to be due to decreased cardiac filling caused by the increased intrathoracic pressure, due in turn to the pulmonary distension and rise of the diaphragm. The cardiac slowing does not occur after bilateral section of the vagus nerves. In no instance out of some 700 decompressions on several different species of laboratory animal did a fatality occur from a single decompression. The injuries which are most likely to occur from explosive decompression are damage to ears and lungs, though rupture of the stomach or intestine may result.

ANOXIA DUE TO ABNORMALITIES OF THE PULMONARY MECHANISM

PNEUMONIA

In *lobar pneumonia* the oxygen saturation of the arterial blood varies in different cases from normal to less than 70 per cent. The signs and symptoms of anoxemia usually appear when the saturation is around 85 per cent. Cerebral symptoms, e.g., sleeplessness and delirium, cyanosis and dyspnea increase with the oxygen desaturation and a saturation of less than 80 per cent is associated with a very high mortality. In 33 cases reported by Stadie with a saturation as low or lower than this only 1 recovered. In lobar pneumonia the carbon dioxide content of the arterial blood is reduced on the average by about 15 per cent (Meakins and Davies). The "alkali reserve", however, is normal or only slightly reduced. The blood pH may in some cases be shifted slightly toward alkalinity

From these findings it appears that there exists a partially compensated alkalosis, induced by the hyperventilation (blowing off of carbon dioxide). In severe cases the excessive elimination of CO_2 may cause a rise in the respiratory quotient to above unity. The increased ventilation of the alveoli also increases the percentage of oxygen in the alveolar air but this fact cannot, as we shall see presently, increase appreciably the oxygen in the arterial blood.

In *bronchopneumonia* a higher degree of oxygen desaturation of the arterial blood is usually present than in the lobar type. The cyanosis may be extreme. There is often *retention* of carbon dioxide when a rise in plasma bicarbonate results, to compensate, in part at least, the gaseous acidosis. The carbon dioxide content of the arterial blood may be 80 volumes per cent or more and the carbon dioxide dissociation curve well above the normal level.

The causes of the anoxia (anoxemia) in pneumonia

The main factors concerned in the production of the anoxemia are (a) the passage of blood through



FIG 34.2 X-ray photograph of lung (injected with barium) from a case of lobar pneumonia (after Gross). *Upper right hand area* Consolidated area—red hepatization. The main vessels are constricted and the finer vascular structure is less dense than in the normal lung. *Lower right hand area* Consolidated area—gray hepatization. The main vessels are patent but the finer vessels have been occluded. *Lower left hand area* Healthy portion of lung except for compensatory congestion, the vessels are dilated.

unaerated (consolidated portions of the lung and (b) shallow breathing. The oxygen unsaturation is not due to any change in the hemoglobin itself (e.g., the formation of methemoglobin) since the blood of pneumonia patients has a normal oxygen capacity. The oxygen dissociation curve at a given carbon dioxide tension is not appreciably different from that of normal persons.

In *lobar pneumonia* during the stages of engorgement and red hepatization the alveoli of the affected portion of the lung are poorly aerated. Mucus blocks the bronchioles and the *air spaces* are filled or their walls coated with exudate. But a large proportion of the vessels of these unaerated regions are still pervious. Consequently blood traversing such areas must remain poorly oxygenated or entirely venous. This blood with a low oxygen saturation and a high carbon dioxide content mixes with blood which has passed through aerated regions. The general arterial blood therefore has its oxygen saturation reduced in proportion to the amount of unsaturated blood with which it is mixed (see also Shunt, p. 433). When the pneumonic area passes into the stage of gray hepatization the vessels of the affected lobe become obliterated to a large extent and the pulmonary blood then passes through aerated regions (fig. 34.2). That is, the arterial blood is no longer vitiated by blood from non-aerated areas. Therefore in a typical case of lobar pneumonia when the disease is confined to a single large area and the breathing is not shallow there is little anoxemia at this stage. If however, bronchopneumonic areas co-exist, the respiratory functions will be affected as described below (See Oxygen Therapy, p. 438).

In *bronchopneumonia*, patches of lung tissue are cut off from their air supply. The fine bronchioles become plugged with mucus, groups of alveoli become filled with exudate and the alveolar walls are edematous and thickened. Yet, obliteration of the vessels to any great extent does not occur. Blood continues to flow through unaerated areas. This blood which is highly venous, mixing with that from aerated alveoli lowers the saturation of the general arterial blood.

SHALLOW BREATHING In pneumonia breathing is frequently very rapid and shallow. Instead of the tidal air being around 500 cc. as in health it may be reduced to 250 cc. or less. It will be recalled that 150 cc. are required to fill the anatomical dead space, therefore only 100 cc. will enter the air sacs of the healthy parts of the lung. We have seen that the expansion of the lungs is

not equal in all its parts (p 350) The alveoli towards the hub of the radiating rays expand less than those near the periphery Those parts, such as the apex, which are indirectly expanded have even in health a tendency to be ventilated less than those which are directly expanded In shallow breathing, these differences are greatly exaggerated Though the volume of the tidal air is only half the normal, the total quantity of air breathed per minute (minute volume) is, as a result of the increased respiratory rate, much greater than normal Since a proportion of the alveoli are very poorly ventilated or not at all, owing to the unequal expansion of the lung, those in other areas tend to be over-ventilated The O_2 tension in the latter is therefore raised But so far as the oxygenation of the blood is concerned the over-ventilation of some alveoli cannot make up for the under-ventilation of others We know that the hemoglobin is nearly saturated already at the ordinary alveolar oxygen tension of 100 mm Hg As already mentioned the dissociation curve of hemoglobin in pneumonia does not differ appreciably from that in health and the most therefore that could be expected from a rise in the alveolar O_2 tension would be an increase of 2 per cent or less in oxygen saturation of the blood traversing over-ventilated regions and a slight increase in the amount of O_2 held in simple solution In other words, the blood flowing through the poorly ventilated parts of the lung will have a low saturation since the O_2 tension is low, while that flowing through the over-ventilated parts will be little above the normal The net result will be a low oxygen saturation of the mixed arterial blood

Matters are different in the case of CO_2 elimination The shape of the CO_2 dissociation curve which shows a progressive slope throughout the entire range of CO_2 tensions is quite unlike that for oxyhemoglobin (fig 34 3) The greater total ventilation results in a lowering of CO_2 tension in the over-ventilated parts of the lung and, consequently, in a "blowing off" of CO_2 from the blood circulating through these regions CO_2 is retained in the blood circulating through poorly ventilated regions In the patient with lobar pneumonia the amount of CO_2 blown off may exceed that retained, the net result will be a lowering of the CO_2 content of the arterial blood In bronchopneumonia a larger proportion of the pulmonary blood circulates through non-aerated areas As a consequence, CO_2 retention is greater, and a normal or

*Bronchopneumonia
CO₂ retained*

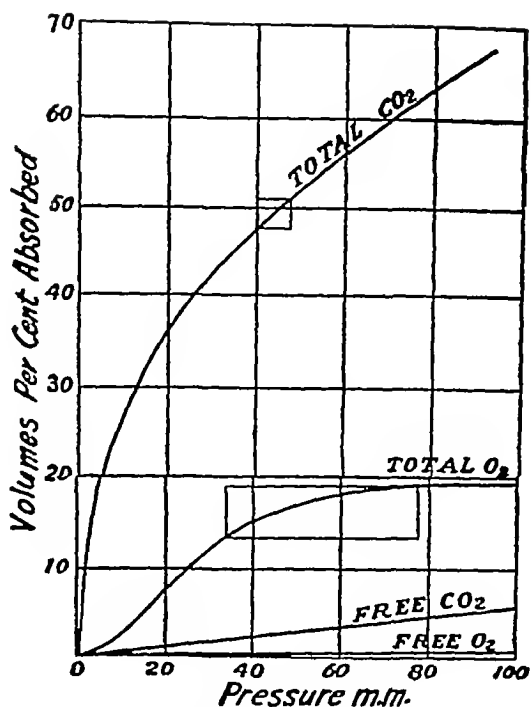


FIG 34 3 Carbon dioxide and oxygen dissociation curves The small rectangles indicate the extent of the variation of the O_2 and CO_2 of the subject's blood when at rest (From L J Henderson)

a higher than normal CO_2 content of the blood is more usual.

Haldane, Meakins and Priestly produced anoxemia in normal persons by having them breathe into an apparatus which reduced the volume of the tidal air As the depth of the breathing became reduced the respirations increased to 100 or more per minute The CO_2 percentage of the alveolar air fell and the O_2 percentage rose The shallow breathing artificially induced in this way resembles that seen in pneumonia and other diseased conditions The administration of oxygen abolished the anoxemia ✓

The cause of shallow breathing in pneumonia Pleuritic pain, by restricting the respiratory excursions, may result in this type of breathing In other instances it appears to be of a reflex nature resulting from the inflammatory process which, through a reduction in the distensibility of the pulmonary tissue, exalts the sensitivity of the afferent vagal endings in the alveolar walls Thus the inspiratory movement is terminated before a full excursion has been completed (see Hering-Breuer reflex, p 407) In support of a reflex origin may be mentioned the experiments of Dunn and of Binger, Brow and Branch who produced this type of breathing in animals by the intravenous injection of potato starch granules These,

acting as small emboli, plugged the pulmonary capillaries. The rapid shallow breathing was immediately abolished by section of the vagi or prevented if the nerves had been cut before the injection. Breathing a mixture rich in carbon dioxide with intact nerves also restored the respiratory rate and depth to normal—the action of carbon dioxide upon the respiratory center itself overcoming the afferent nervous influence. Also, in pneumonia it has been found that oxygen inhalations, even though they may restore the oxygen saturation of the arterial blood to normal, do not necessarily abolish the shallow breathing—further evidence for the existence of a nervous element in the production of this type of breathing. Shallow breathing may also result from other diseases involving the alveoli, e.g., inflammation by irritant gases (phosgene and chlorine), pulmonary edema, miliary tuberculosis and pulmonary emboli, which would be expected to stimulate afferent nerve endings. It also occurs in certain nervous states, hysteria, certain forms of neurasthenia, and sometimes in encephalitis lethargica, but is not seen in lesions involving the bronchi or bronchioles alone, e.g., bronchitis and asthma.

Severe anoxemia, however produced, tends itself through its damaging effect upon the respiratory center to induce rapid shallow breathing, and in any event will exaggerate this type of breathing, since it also tends to increase the sensitivity of the Hering-Breuer reflex. Thus a vicious circle—shallow breathing inducing anoxemia and the latter reacting to enhance the former—is set up. Shallow breathing induces one of the gravest types of anoxemia, since the cardiovascular system, as well as suffering from oxygen want, is seriously affected by the excessive loss of carbon dioxide (see p. 292). Owing to the narrowed state of the cutaneous vessels the cyanosis is of the pale leaden-gray type (p. 437) and if the anoxia is not relieved, failing circulation adds to the oxygen want (anoxia of the stagnant type). The apnea would also tend to magnify the oxygen want of the tissues for, as we have seen, the hemoglobin gives up its oxygen less readily at low carbon dioxide tensions.

ASTHMA

This is a paroxysmal disease in which acute oxygen want is caused by a spasm of the smooth muscles of the finer bronchioles. Edema of the bronchiolar mucosa is probably also present. The alveoli are poorly ventilated and some may be

completely cut off from their air supply. The high percentage of carbon dioxide and low-percentage of oxygen in the alveolar air result in a low oxygen saturation of the arterial blood and the retention of carbon dioxide. The gaseous acidosis is met by the excretion of a highly acid urine and a rise in the "alkali reserve." An intense plum-colored cyanosis may develop. The continued stimulating effect of oxygen want and carbon dioxide excess upon the respiratory mechanisms causes severe dyspnea.

Difficulty is experienced both in inspiration and expiration but since there is a natural tendency for the bronchioles to narrow during expiration and dilate during inspiration the greatest respiratory effort is exerted during expiration. The respiratory muscles contract with great force and the accessory muscles of respiration are brought into play. The expiratory muscles compress the chest and the abdominal muscles contract in the attempt to squeeze the air from the lungs. The intrapulmonary pressure is greatly elevated and the air escapes through the constricted tubes with a distinct wheezing sound. Owing to the difficulty and prolongation of the expiratory phase normal deflation of the lungs cannot occur before the next inspiration ensues. The lungs, therefore, remain almost maximally expanded even at the end of expiration. That is, during the asthmatic paroxysm a very large volume of residual air is present in the lungs (fig. 344). The volume of the tidal air is greatly reduced and corresponds to the vital capacity at the moment. Since the subject is already exerting the greatest inspiratory and expiratory efforts of which he is capable there can be no supplemental or complemental air. The changes in volume of the over-distended lung are small and not commensurate with the excursions of the thoracic walls. As a consequence, the high value of the intrathoracic negative pressure induced during inspiration causes the structures at the root of the neck to be drawn toward the thoracic cavity to take up the space which the lungs are unable to fill. During expiration the veins of the neck and face become engorged. The restricted movements of the lung also greatly reduce the effect of mechanical mixing upon the lung air, the slower process of diffusion being depended upon to a larger extent for the freshening of the alveolar air.

The sputum in asthma contains spirals of delicate fibrils (Curschmann's spirals) formed of bronchiolar secretion, diamond shaped crystals (Charcot-Leyden

crystals) and eosinophil cells. A substance possessing a histamine-like action has been demonstrated in the sputum by Knott. The blood in asthma shows a great increase in the eosinophil cells (eosinophilia). The effect of the asthmatic paroxysm upon the heart is that of partial asphyxia. A-V conduction may be depressed as shown by some lengthening of the P-R interval in the electrocardiogram or premature beats may occur. The attack appears to cause no permanent damage to the cardiovascular system.

Causation. Asthma often shows a strong hereditary tendency. The bronchiolar spasm may be (a) of a reflex nature and due to the stimulation of hypersensitive afferent vagal endings in the larynx, or of trigeminal fibers by some nasal abnormality, the bronchoconstrictor impulses travel via efferent vagal fibers, (b) an allergic phenomenon, i.e., the result of sensitization to some foreign protein. This is the most common cause of the condition. It is then frequently associated with other allergic conditions, e.g., hay fever, urticaria or eczema either in the patient himself or in members of his family. The foreign protein may be inhaled. Pollens of various grasses and flowers, the dandruff of animals, e.g., horse, cat or dog, or feathers are among the most common excitants, the exciting cause may be some kind of food or the protein of bacteria within the respiratory tract itself may be responsible. The sensitivity of certain individuals to foreign proteins presents many features resembling those of anaphylactic shock in animals. The two conditions are probably in some way closely related. Sudden death may result from the injection of horse serum (e.g., diphtheria-antitoxin or anti-tetanic serum) into an asthmatic subject. A guinea pig when injected with a protein to which it has previously been sensitized dies rapidly from anaphylactic shock. The bronchiolar muscle is strongly contracted. The air is trapped so that the lungs are maximally distended and do not collapse when the thorax is opened. Even when the pulmonary tissue is deeply incised the air does not escape from the distended lung. This manifestation of anaphylaxis is associated with, perhaps due to, the liberation of histamine (see ch. 27). It is well known that anaphylaxis and histamine administration produce almost identical effects in the guinea pig (fig. 34 5). Histamine producing bacteria have been reported in the bronchial secretions of asthmatics.

The treatment of asthma resolves itself into the relief of the paroxysm and the removal of the

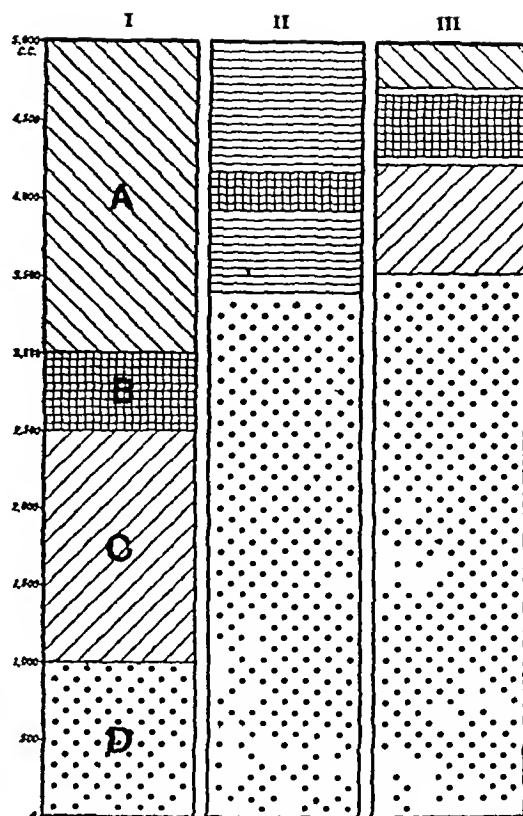


FIG 34 4 Diagram showing subdivisions of the lung air in asthma (II) and emphysema (III) compared with the normal (I). A, complementary air, B, tidal air, C, supplemental (reserve) air, D, residual air. The horizontal lines above and below the area representing tidal air in II and III indicate the extent of the respiratory movements. (After Coke.)

37

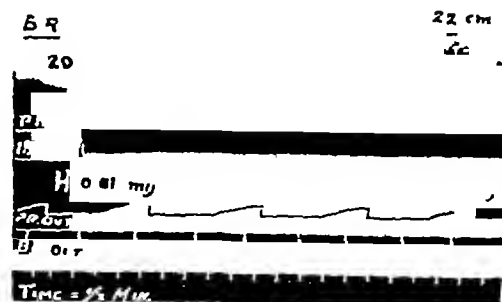


FIG 34 5 Guinea-pig. Reaction of bronchial muscle to 0.01 mg histamine. (After Thornton)

underlying cause Adrenaline or ephedrine acts by inhibiting the bronchiolar muscle during the attack Atropine paralyzes the broncho-constrictor (vagal) fibers Of the three drugs, adrenaline is the most effective In the allergic form of the disease every effort is made to identify the offending protein Antihistamine drugs (p 296) have been employed

CHRONIC EMPHYSEMA

(Greek *em* + *physema*, a blowing)

The lungs in emphysema are in a state of extreme distention as a result of the enlargement of the air sacs The latter, however, show fewer alveoli in their walls owing to the atrophy of the inter-alveolar septa Contiguous air sacs within a lobule coalesce or even adjacent lobules may fuse to form large air spaces For this reason the total respiratory surface is reduced The alveolar and capillary walls become thickened and the interstitial pulmonary tissue increased Many capillaries become occluded The pulmonary elastic tissue is reduced in amount so that the lungs, when removed from the thorax, do not collapse normally but remain in an over expanded state The peripheral lobules, which in health expand to the greatest extent, are those mainly affected in emphysema, the enlarged lobules appearing as blebs upon the surface of the lung

The chest is what is known as barrel-shaped The ribs are more horizontal than normally, the thoracic spine is bowed backwards (kyphosis) so that the anteroposterior diameter of the chest is as great as or exceeds the transverse The position of the chest is one of nearly full inspiration The mid-position of the diaphragm is at a much lower level than usual and its excursions above and below this level are restricted The respiration is therefore mainly costal In some instances the diaphragm is practically fixed or indeed may be drawn up during inspiration (paradoxical movement) As in the paroxysm of asthma the residual air is 2 or 3 times the normal and the complementary air is reduced The tidal air is normal or only moderately reduced and the vital capacity (fig 34.4) is lowered by from 20 to 60 per cent Owing to the loss of the elasticity of the lung, expiration is no longer simply a passive movement but is aided by a forcible contraction of the expiratory muscles When the patient is asked to make a deep inspiration he does not expel all the air during the next expiration. A series of respirations occur before

the chest returns to its original size Furthermore, he cannot expel nearly as large a volume of reserve (supplemental) air immediately after a deep inspiration These phenomena are due to the fact that the inelastic lungs have been overstretched and are brought back to their original volume with difficulty

In emphysema there are anoxemia and retention of carbon dioxide The oxygen saturation of the arterial blood runs from a little below normal to around 75 per cent or even less The carbon dioxide tension in the alveolar air in well marked cases is from 50 to over 60 mm Hg (7 to 8 per cent) and the carbon dioxide content of the arterial blood correspondingly high

Owing to the obliteration of many capillaries in the lungs, the resistance in the pulmonary circuit is increased, a greater burden being thus put upon the right ventricle

The cause of the impaired gaseous exchange is not altogether clear Thickening of the alveolar and capillary walls and the obliteration of capillaries have been considered to be a factor Yet if this were so one would not expect the retention of carbon dioxide which, owing to its greater solubility (30 times that of oxygen) has a much higher rate of diffusion through the pulmonary membrane, to be so much more pronounced than the anoxia The sharp rebound at the end of inspiration which occurs in the healthy lung causes mechanical mixing of the lung air and is an important factor in the efficient ventilation of the alveoli The absence of this effect in the emphysematous lung and its greater dependence, in consequence, upon the slower process of diffusion is probably an important factor leading to the defective aeration of the blood According to Christie, the impaired gaseous exchange is due mainly to the fact that, as a result of the loss of elasticity, the intrathoracic pressure is not distributed evenly throughout the lung As a consequence, the outlying alveoli which are largely functionless with obliterated vessels are ventilated to a greater extent than the relatively healthy ones more centrally placed Owing to the shapes of the respective dissociation curves such underventilation of the functioning alveoli will tend to have a greater effect in preventing the elimination of CO₂ than in interfering with the absorption of oxygen Another factor which is probably of importance is the slower rate of diffusion of CO₂, owing to its larger molecule, in the alveolar air

The red cell count, hemoglobin percentage, and

consequently the oxygen capacity of the blood, are increased above the normal in emphysema. The cyanosis (p 437) is often pronounced, yet the patient's dyspnea is less than might be expected from his color, often startling, and from the carbon dioxide retention which exists. This is explained by the well established fact that in emphysema the respiratory center is relatively insensitive to carbon dioxide. A normal person when breathing a carbon dioxide rich mixture (8 per cent) increases his pulmonary ventilation by 300 per cent or more, the breathing of the emphysematous patient on the other hand shows relatively little change as a result of breathing a much stronger mixture (see fig 34 6)

Causation Two factors are concerned in the production of emphysema (a) reduction in the elastic tissue of the lung and (b) increased distention of the alveolar spaces

It is very questionable whether, in the absence of some abnormality of the lung tissue itself, emphysema can result from increased intrapulmonary pressure, such as occurs in those following certain occupations, e g, glass blowers and the players of wind instruments. The study of groups of men following such occupations does not indicate that emphysema is produced in this way. Emphysema has, however, been induced in animals by stenosis of the trachea or bronchi or by the insertion of a valved apparatus into the trachea which allowed the free ingress of air but offered an obstruction to expiration. It is probable that in these instances the persistently high intraalveolar pressure by stretching the lung structures during inspiration first caused atrophy of the elastic tissue. During inspiration the air spaces are dilated by the negative pressure upon their outer surfaces. The trapped inspired air causes an ever-increasing distending pressure to be exerted upon the alveolar walls. Inasmuch as the intrathoracic pressure is "less negative" during expiration, a reduction rather than an enlargement of alveolar capacity would result at this time (see p 168). Asthma and chronic bronchitis, which frequently are forerunners of emphysema also, probably, exert their damaging effect upon the alveolar structure during the inspiratory phase. In the former condition the spasm of the bronchiolar muscle exerts a valve-like action (see p 359). In chronic bronchitis mucous plugs would act similarly. Coughing, it has been supposed, places a strain upon the alveolar walls. But during the phase of coughing when the glottis is closed the alveolar walls are supported. When the glottis opens, the air escapes from the alveoli if the obstructing material has been dislodged, and no strain upon the alveolar wall would result. If, however, the air remains entrapped its sudden re-expansion (re-bound), as the pressure in the surrounding pulmonary tissue falls at the

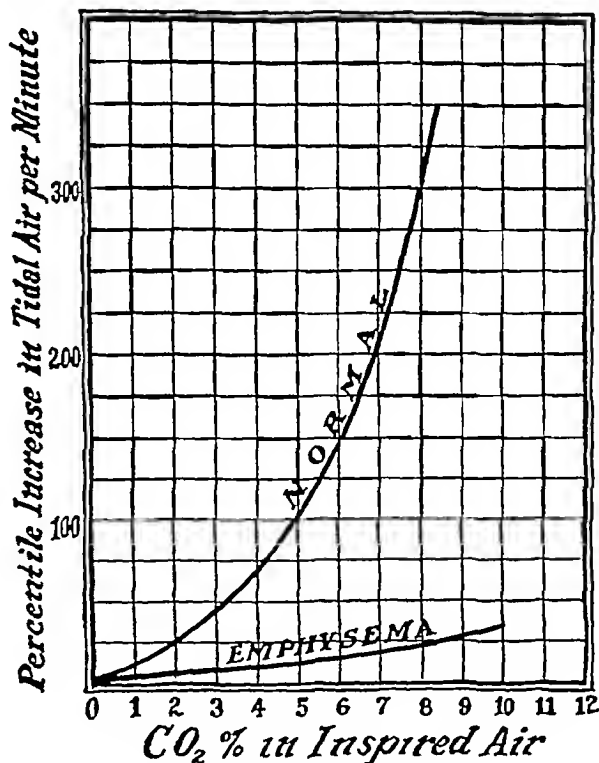


FIG 34 6 Chart showing the percentile increase in tidal air per minute as the percentage of inspired carbon dioxide is raised. Note that when the normal subject inspires air containing 8 per cent carbon dioxide, the tidal air is increased about 300 per cent, whereas in the emphysematous subject it is increased only about 25 per cent (After R W Scott.)

end of a cough, may injure the alveolar membrane and start the emphysematous process.

What may be termed a physiological or compensatory emphysema occurs when part of the pulmonary surface is reduced as by the collapse of a part or the whole of one lung. This is more in the nature of an hypertrophy. A similar enlargement of both lungs occurs at high altitudes.

A type of emphysema also occurs in the elderly—sensile or postural emphysema—and is secondary to the change in the shape of the thorax which becomes more barrel-shaped and increased in capacity. The lungs enlarge to fill the increased space. This condition is associated with few symptoms. There is little reduction in the vital capacity and the oxygen saturation of the arterial blood is practically normal.

COLLAPSE OF THE LUNGS OR ATELECTASIS

(Gk *ateles*, incomplete, *ektasis*, distention)

Any condition which lowers the pressure (intrapulmonary) within the alveoli or increases the pressure upon the lung surface, i e, reduces the "negative" intrapleural pressure, may lead to collapse of the lung. Thus pleural effusions, pneumothorax,

tumors, etc., pressing from without or the isolation of the alveoli from their air supply by the obstruction of a bronchus will therefore result in collapse of the lung or of the portion of the lung affected. Atelectasis is also the term applied to the condition in the newborn in which, as a result of the blockage of a bronchus or of a group of bronchioles by mucous secretion, or owing to weak inspiratory movements, a portion of the lung fails to become distended with air.

When a bronchus or bronchiole in a previously distended lung is obstructed the imprisoned air soon becomes absorbed from the affected alveoli. Collapse of the air sacs cannot take place until this has occurred. Absorption is brought about in the following way, as pointed out by Henderson. The air in the isolated alveoli has a total pressure of 760 mm. Hg (p. 357). The partial pressures are in round numbers, O, 100 mm., N, 570 mm. CO₂, 40 mm. and aqueous vapor, 47 mm. In the venous blood the total pressure is 703 mm., the nitrogen and aqueous vapor being the same as that of the alveolar air, but the partial pressure of oxygen is only 40 mm. and of carbon dioxide 46 mm. An interchange of the latter gases therefore occurs between the alveolar air and the venous blood. It might be thought that the imprisoned air would then be in equilibrium with the blood and no further absorption occur. But the alveolar air loses more oxygen than it gains carbon dioxide whereupon its volume is reduced. The atmosphere acting upon the body surface and through the yielding soft tissues compresses the air so as to maintain its total pressure practically constant at 760 mm. As a result of the absorption of oxygen the percentage and consequently the partial pressure of carbon dioxide and of nitrogen are increased. These gases now diffuse into the blood. The volume of the alveolar air is further reduced thereby but its total pressure still remains unaltered. The percentage and therefore the partial pressure of oxygen rises and more of this gas passes into the blood. The process continues in this manner until no air remains, and the walls of the original space are ultimately approximated by the pressure of the atmosphere. Air is absorbed from the pleural cavity or from any other closed cavity of the body in the same way (Henderson and Henderson). The absorption of the air confined within the pleural cavity permits the lung, if the air pressure had caused its collapse, to re-expand (see also p. 358).

The collapse and shrinkage of the lung which results from blockage of a bronchus increases the intrathoracic negative pressure, since the closed thoracic box is less completely filled. The diaphragm is therefore drawn upwards and uncollapsed portions of the lungs are expanded to a greater extent (compensatory emphysema) in order to fill the unoccupied space.

SHORT

In the fetus the vessels of the airless lungs are by-passed by the greater part of the blood brought to the heart. That is to say, most of the blood is shunted to the arterial side through channels which normally close at, or shortly after, birth. It is clear that if one or other of these channels persist after birth or if the interventricular septum is defective so that a large part of the venous blood does not traverse the lungs but is short-circuited from the right to the left heart or directly into the aorta, the oxygen saturation of the arterial blood will be seriously reduced. Anorexia, evidenced by cyanosis (p. 435) and dyspnea (p. 414) will result (Fig. 347). These will be especially pronounced upon exertion since the unsaturation of the pulsating venous blood will be thereby increased.

(*Temporal circulation*—Before the congenital abnormalities are described the fetal circulation and the circulation readjustments which occur at birth, or shortly thereafter, will be briefly reviewed.

In the fetus the oxygenated blood from the placenta is carried on the umbilical vein to the liver, where it separates into two streams, one of which is distributed to the left $\frac{1}{3}$ or so of the liver; the remainder passes directly via the ductus venosus into the inferior vena cava, thus mixing with blood returning from the lower part of the body. The portal vein supplies the rest of the liver. This portion of the blood from the umbilical vein which has traversed the vessels of the liver is also delivered by the hepatic veins into the inferior vena cava. The hepatic veins are a number of small vessels and two much larger ones, the latter form a junction with the ductus venosus just before the latter joins the inferior vena cava. The blood in the vena cava beyond this point is, therefore, partly oxygenated (from placenta via umbilical vein and ductus venosus) and partly reduced, having traversed the tissues of the lower limbs, intestines and liver. The caval blood, upon reaching the right auricle, separates into two unequal parts. The much larger stream passes directly through the foramen ovale (see structure) into the left auricle, the smaller stream enters the right ventricle. The blood which has reached the left auricle, after being joined by a small

but by no means inconsiderable volume of blood returned from the pulmonary tissue, flows into the left ventricle, and is discharged into the aorta and distributed to the brain by the brachio-cephalic artery, and to the myocardium by the coronary arteries. The remainder flows down the aorta.

The blood returned from the head and entering the right auricle by the superior vena cava, together with the small fraction of blood of the inferior vena cava which was not directed through the foramen ovale, enters the right ventricle and is ejected into the pulmonary artery. A smaller part, yet, as shown by Franklin and his associates, more than has been generally supposed is conveyed to the lungs by two branches of the latter vessels. The main stream, discharged from the right ventricle, passes by a wide vessel, the *ductus arteriosus*, into the aorta, and mixing with the blood coming from the left ventricle, is distributed throughout the body. That portion which is supplied to the lower limbs is returned by two vessels—the *umbilical arteries*—to the placenta.

During birth, or very shortly after, the three umbilical vessels close (even though the cord has not been tied or broken) by the contraction of smooth muscle in the vascular walls. The flow through the ductus venosus ceases as a result of contraction of a smooth muscle sphincter at the point where it leaves the umbilical vein. This occurs in the sheep fetus, according to Franklin and his associates, in from 5 to 25 minutes after the umbilical vessels have closed. Within 5 minutes or so after the commencement of breathing the foramen ovale closes, and a little later (apparently never before the obliteration of the foramen ovale) the ductus arteriosus becomes occluded. All the blood from the right side of the heart is normally from now on directed through the lungs. Franklin and his associates found, however, in the case of the new-born lamb, that should the general condition of the latter seriously deteriorate, the ductus arteriosus may open again. The mechanism of closure of the ductus appears to be initiated in some way by the action of the oxygenated blood upon the vascular muscle and not through a nervous mechanism. In the lamb this channel has never been observed to close before occlusion of the umbilical vessels.

Congenital cardiac defects

/ Patent ductus arteriosus, or foramen ovale inter-ventricular septal defects, stenosis of the pulmonary artery and coarctation of the aorta (narrowing of aorta in the neighborhood of the ductus arteriosus) are among the congenital defects most commonly met with.

The development of the direct Fick method (catheterization of the right side of the heart) for determining the cardiac output has provided means of the greatest value in diagnosing congenital defects of the circulation. It is now possible to obtain samples of

blood for gas analysis from the right auricle or ventricle, or from the pulmonary artery (See fig 34 8)

Patent foramen ovale or other defect in the interauricular septum In patency of the foramen ovale, the opening is guarded by a valve-like membrane which prevents blood from passing from the left to the right auricle, and none passes in the opposite direction unless the pressure in the right auricle is high. With other defects of the inter-auricular septum, a shunt from right to left with cyanosis may occur, but if there is mitral stenosis and a

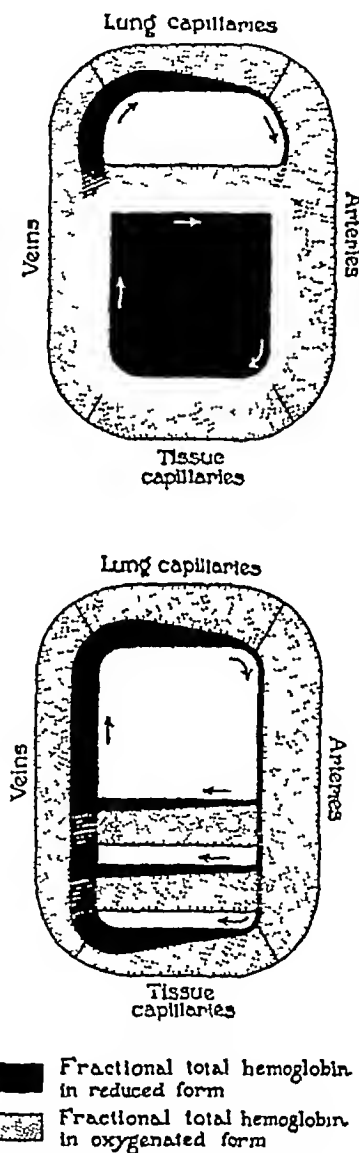


FIG 34 7 Diagram (upper) showing the proportion of oxyhemoglobin to reduced hemoglobin in different parts of the circulation in an instance where a portion of the blood passes through un-aerated channels (shunt) from the venous to the arterial system. Lower diagram represents a case in which the oxygen unsaturation of the blood is abnormally high in a part of the peripheral capillaries but normal in the arterial blood (stagnant type of anoxia) (After Lundsgaard and Van Slyke)

large defect (*Lutembacher's syndrome*) the high pressure in the left auricle causes a shunt from left to right

Patent ductus arteriosus Blood enters the pulmonary system from the arterial side, the oxygen content in the pulmonary artery is increased, but is normal in the right auricle and ventricle

If a defect of the *interventricular septum* exists alone the oxygen content of the blood in the right ventricle and pulmonary artery is increased but not of that in the right auricle

Neither of these last two defects, if it exists alone, will cause cyanosis, since the shunt is from left to right (arteriovenous shunt) Rarely, however, in the case of a patent ductus arteriosus, the direction of flow is reversed temporarily at certain times (during crying or coughing), blood passing from the pulmonary artery into the aorta.

Often the abnormalities are multiple For example, a patent foramen ovale, together with stenosis of the pulmonary artery, results in the passage of venous blood to the arterial side, i.e., through the auricular defect, slight or moderate cyanosis is seen The combination of congenital defects, originally described by John Hunter and later by Fallot, and now generally known as the *tetralogy of Fallot*, consists of, (a) an *inter-ventricular septal defect*, (b) *narrowing of the pulmonary artery*, or stenosis of its orifice, and (c) *dextro-position of the aorta*, the orifice of this vessel being shifted to the right to straddle the septal defect. (d) The fourth element making up the tetrad is *right ventricular hypertrophy* The blood flow through the lungs and the pulmonary pressure are reduced Owing to the high pressure developed in the *right ventricle* blood is driven through the defect into the overriding aorta. If the pulmonary stenosis is very severe, the ductus arteriosus must be patent in order to provide

sufficient blood flow through the lungs to support life. Cyanosis and dyspnea are usually pronounced. Blalock and Taussig have devised an operation for the correction of the chief effects of these conditions It consists of anastomosing a large branch of the aortic arch (usually carotid or subclavian) to the right or left *pulmonary artery*

Another combination of defects has three features of the tetralogy of Fallot, namely, *defect of the intra-ventricular septum*, *dextroposition of the aorta* and *right ventricular hypertrophy*, but pulmonary stenosis is absent. It is known as *Eisenmenger's complex* The shunt is from left to right in early life, but later, when as a result of the large volume of blood entering the right ventricle from the left side as well as through the normal channels this chamber enlarges and hypertrophies, the high pressure developed may force venous blood into the overriding aorta, cyanosis then appears

In those congenital cardiac anomalies showing cyanosis, the arterial oxygen saturation is found to be, during rest, from 45 per cent in severer degrees of cyanosis to 90 per cent in the milder cases The red cell count, however, is likely to be increased and the hemoglobin concentration may be as high as 26 grams per cent. The arterial oxygen content, i.e., the volumes of oxygen per 100 cc. of blood, in those with the higher hemoglobin concentration is, consequently, much higher than normal and may reach a value of 29 volumes per cent, as against the normal maximum of 22 volumes per cent (see Cyanosis, p. 435) Exercise, which in normal subjects causes little or no change in arterial oxygen saturation, produces a profound fall in the congenital heart cases, amounting in some to a reduction of over 30 per cent.

CARBON MONOXIDE AND CYANIDE POISONING

CARBON MONOXIDE POISONING

References to carbon monoxide poisoning are contained in the earliest medical writings This gas was used by the Greeks and Romans for the execution of criminals and as a means of committing suicide Carbon monoxide is today the most important gaseous poison against which physicians have to contend During times of peace it accounts for more deaths than all the other gases combined Carbon monoxide combines with the hemoglobin of the blood and thus renders it unavailable for oxygen carriage The affinity of hemoglobin for carbon monoxide is approximately 300 times its affinity for oxygen Therefore, when the atmosphere contains only a very small percentage of CO the hemoglobin takes up the poisonous gas and in so doing prevents its combining with an equivalent volume of oxygen

Recovery from carbon monoxide poisoning is

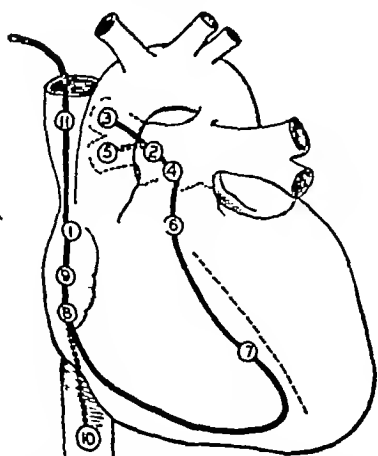


FIG 34 8 Numbers from 1 to 11 indicate the positions from which blood samples may be taken by means of the intracardiac catheter (After Groom and associates)

usually complete when the exposure has not been too long or the concentration too high. It is important that the carbon monoxide hemoglobin should be broken up as soon as possible since injury to the tissues is produced by the anoxia. It is possible to displace the carbon monoxide by oxygen if the tension of oxygen is sufficiently high and that of carbon monoxide low. Mixtures with a high percentage of oxygen and from 6 to 7 per cent carbon dioxide combined with artificial respiration are used in the treatment of CO poisoning. The carbon dioxide, as well as acting as a respiratory stimulant, reduces the affinity of hemoglobin for carbon monoxide. Intravenous injections of methylene blue have been advocated but are valueless. Carbon monoxide, as well as its action in displacing oxygen from hemoglobin, has histotoxic properties, inhibiting the tissue respiratory enzymes. This fact has no practical bearing, however, upon carbon monoxide poisoning in man, for the concentrations at which such action occurs is many times greater than could ever occur in the body.

CYANIDE POISONING

The mechanism of production of anoxia by cyanide is quite different from the action of carbon monoxide. No interference with oxygen carriage is caused by cyanide but there is inhibition of tissue respiration. Cyanide inhibits the action of cytochrome oxidase, carbonic anhydrase and probably of other enzyme systems (see ch. 32 and p. 420). It follows from this that any antidote for cyanide poisoning must have one of two actions. Either the cyanide must be removed or detoxified or the inactivated catalyst must be replaced. Methylene blue does act as a catalyst for certain biological oxidations and this led to its use as an antidote for cyanide poisoning. When methylene blue is given to the intact animal there is a marked rise in body temperature which is due to increased metabolism. Methylene blue permits inhibiting effects of cyanide on cellular oxidations. This fact has been established from investigation on a great number of tissues. Sahlén in 1926 provided the first experimental demonstration that methylene blue antagonizes the action of cyanide in the intact animal. He used rats and the observation has been confirmed on other animals—dogs, rabbits, mice. The evidence suggests that methylene blue acts by removing the cyanide from tissue. Methylene blue and cyanide do not combine directly but methylene blue forms methemoglobin and the methemoglobin combines with the cyanide to form cyanmethemoglobin. The cyanmethemoglobin is relatively non-toxic and is broken down slowly, the detoxification probably being brought about by conversion of the cyanide to thiocyanate (Smith and Mal-

colm). Other substances which form methemoglobin (Hug, Wendel) such as sodium nitrite, amul nitrite, pyrogallol, etc., are also effective in the treatment of cyanide poisoning. Such methods of course are limited by the quantity of hemoglobin that can safely be converted to methemoglobin.

A number of sulphur compounds have been found effective in cyanide poisoning. Chen, Rose and Clowes showed that sodium thiosulphate and sodium tetra-thionate may protect dogs against as much as three lethal doses of cyanide.

CYANOSIS

(Greek, *cyanos*, blue)

Cyanosis may be defined as the diffuse, dusky or bluish color of the skin and mucous membranes caused by the presence in the blood of the superficial capillaries (subpapillary venous plexus, see p. 314) or reduced hemoglobin above a certain definite amount.³ Cyanosis is seen in the anoxic and stagnant types of anoxia but not in the anemic or histotoxic forms. It seems scarcely necessary to state that the retention of carbon dioxide in the blood has no direct effect upon the production of cyanosis. The blue color of the skin depends fundamentally upon the absolute amount of reduced hemoglobin in the capillary blood and not upon the relative proportions of reduced hemoglobin and oxyhemoglobin. For example, in anemia the hemoglobin content of the blood may be only 20 per cent of the normal. In the capillary blood all of this could be in the reduced form yet cyanosis would not result, since the absolute amount of reduced hemoglobin (i.e., "blue pigment") would be insufficient to produce any blue discoloration. On the other hand, in polycythemia the hemoglobin may be 100 per cent above normal. Cyanosis will occur when the hemoglobin of the capillary blood is only 20 per cent reduced, for the absolute concentration of reduced hemoglobin will then be raised to threshold value. The greater quantity of the bright-colored oxyhemoglobin present exerts little or no influence, that is, it does not, as might

³ The presence of abnormal compounds of hemoglobin, e.g., methemoglobin and sulfohemoglobin, resulting from the action of various toxic substances, causes a type of cyanosis (enterogenous cyanosis), but these will not be considered here (see p. 59). Cyanosis may result from anoxia of either the anoxic or stagnant type. It obviously cannot occur in the anemic type, which is due essentially to a low hemoglobin concentration, in the histotoxic type in which the hemoglobin gives up less of its oxygen store than in health, nor in the anoxic and stagnant types if a severe grade of anemia exists.

be expected, tend to neutralize the color effect of the reduced hemoglobin

Normal blood contains about 15 grams of hemoglobin per 100 cc. Lundsgaard found that the capillary blood must contain approximately 5 grams of reduced hemoglobin per 100 cc. before cyanosis will appear. When fully saturated (20 volumes per cent) $\frac{1}{2}$ gram of hemoglobin will take up 1 cc. of oxygen. Five grams of hemoglobin, therefore, hold about 6.7 cc. of oxygen, and 5 grams of reduced hemoglobin are formed when the blood contains 13.3 volumes per cent of oxygen. Cyanosis may therefore be expected to appear when the blood in the capillaries is on the average around 7 volumes per cent unsaturated. As a result of certain modifying factors (p. 437) the precise level of capillary unsaturation at which cyanosis makes its appearance varies in different cases between 6 and 7 volumes of oxygen per cent.

The oxygen unsaturation of the capillary blood does not of course occur abruptly at the arterial end but is progressive from point to point along the course of the vessel. The loss of oxygen may be uniform from the arterial to the venous end of the capillary as shown in Curve I, figure 34.9, or the desaturation may occur mainly toward the venous end (Curve II) when the capillary blood would approximate arterial blood in its content of reduced hemoglobin. Under other circumstances the greatest oxygen loss may occur toward the arterial end (Curve III) when the unsaturation of the capillary blood throughout would approach that of venous blood. It is difficult to obtain data from which the true curve may be drawn. The simplest of these curves (Curve I) is assumed and the average unsaturation of the capillary blood is taken as midway between that of arterial and venous bloods respectively. Thus

$$\frac{1}{2}(A + V) = C$$

where A and V represent the unsaturation in volumes per cent of arterial and venous bloods

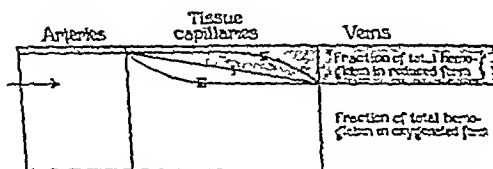


FIG. 34.9 Diagram showing hypothetical variations in the shape of the reduction curve of oxyhemoglobin during the passage of the blood through tissue capillaries. (After Lundsgaard and Van Slyke.)

respectively and C the average unsaturation of the capillary blood

For example, in a normal person the arterial unsaturation is 0.5 volume per cent (19.5 volumes per cent saturation), another 5 volumes per cent are given up in the capillary, the venous unsaturation is therefore 5.5 volumes per cent. So

$$\frac{1}{2}(0.5 + 5.5) = 3.0 \text{ volumes per cent average unsaturation of capillary blood}$$

$$(3.0 \times 0.75) = 2.2 \text{ gram reduced Hb}$$

We have seen that the average unsaturation of the capillary blood must be between 6 and 7 volumes per cent (approximately 5 grams of reduced Hb) before cyanosis appears. This degree of unsaturation of the capillary blood may be brought about either by an increase in the arterial unsaturation (anoxic type of anoxia) or as a result of a greater amount of oxygen being lost from the blood in its passage through the capillaries (stagnant type of anoxia). That is, by an increase in the venous unsaturation (V) alone. In order to produce an average unsaturation of from 6 to 7 volumes per cent in the capillary blood the arterial unsaturation would need to be from 4 to 4.5 volumes per cent, or the unsaturation of the venous blood—that in the arteries being normal—11 to 13 volumes per cent. Meakins and Davies found that when the veins of the arm were obstructed (stagnant type of anoxia) in normal individuals cyanosis was just detectable when the venous blood reached an unsaturation of 11.4 volumes per cent.

These facts may be clarified by examples. If the blood as it leaves the lungs contains only 15.5 volumes per cent of oxygen, i.e., if it has an unsaturation of 4.5 volumes per cent and the tissues abstract the usual quantity of oxygen, namely 5 volumes per cent, the unsaturation of the venous blood will be 9.5 volumes per cent and that of the capillary blood

$$\frac{1}{2}(4.5 + 9.5) = 7 \text{ volumes per cent}$$

$$(7 \times 0.75 = 5.2 \text{ grams Hb})$$

If on the other hand the arterial unsaturation is around the normal value of 0.5 volume per cent but as a result of slowing of the circulation each unit volume of the blood gives up a greater quantity of oxygen, and the venous unsaturation is increased to say 13 volumes per cent, the average unsaturation of the capillary blood will be

$$\frac{1}{2}(0.5 + 13) = 6.7 \text{ volumes per cent}$$

In either of these instances slight cyanosis would be expected to appear

Let us now consider what occurs when the hemoglobin percentage is above or below the normal value

In *anemia* the hemoglobin is below normal and the oxygen capacity of the blood is correspondingly lowered. If the hemoglobin content be only 30 per cent of the normal, the volumes per cent of oxygen in the arterial blood, though the latter be fully saturated, will be only about 6 volumes per cent, i.e., the quantity of hemoglobin in 100 cc of blood is less than 5 grams. It is obvious that such a person could not become cyanosed even if, as a result of defective oxygenation of the blood or of slowing of the circulation, all his hemoglobin were in the reduced state. A patient whose hemoglobin was, say, 60 per cent (oxygen capacity 12 volumes per cent) would, like one with a normal hemoglobin content, become cyanosed when his arterial blood reached an unsaturation of about 4.5 volumes per cent, that is when his capillary blood had an average unsaturation of around 7 volumes per cent. But the oxygen want of the anemic subject would be greater than that of the subject with a normal hemoglobin content since in the case of the former 4.5 volumes per cent constitute nearly 40 per cent of the oxygen capacity of his blood. In a less enlightened age when bleeding was resorted to for the relief of cyanosis, the brilliant success of that procedure is not to be wondered at—nor that the patient died. In the stagnant type of anoxia the unsaturation of the venous blood of a subject with 60 per cent hemoglobin content could not reach the value necessary to produce cyanosis until all the hemoglobin was in the reduced state. Thus $\frac{1}{2}(1 + 12) = 6.5$ volumes per cent unsaturation of the capillary blood.

In *polycythemia*, in which the hemoglobin content is, say, double the normal (oxygen capacity 40 volumes per cent), cyanosis occurs in very mild degrees of anoxemia. At an arterial unsaturation of 4.5 volumes per cent the polycythemic subject would have the same degree of cyanosis as an ordinary person with this quantity (5 grams) of reduced hemoglobin in his blood. But in the polycythemic subject with a hemoglobin content of 200 per cent, 4.5 volumes per cent is only 11 per cent of the oxygen capacity of his blood, in the person with a normal hemoglobin content it constitutes 22.5 per cent. The oxygen want in the former would be relatively slight as compared with

that of the latter. Or put in another way and including the anemic subject—if the anoxemia were of the same degree in each, the anemic person would show little or no cyanosis, the polycythemic, a cyanosis of high degree and the normal subject a color intermediate in intensity. In a person with an abnormally high blood count a degree of slowing of the peripheral blood flow which would be without effect upon one possessing a normal hemoglobin content will result in cyanosis. Thus in regions such as the face, ear lobes and hands, where the cutaneous vessels are well filled with blood, cyanosis is readily induced in the polycythemic subject by exposure to cold. In certain conditions associated with arterial anoxemia, e.g., emphysema (p. 430) and congenital heart disease the red cell count is increased and the cyanosis, in consequence, enhanced. It will be evident from these examples that cyanosis is not a reliable guide for the detection of anoxia.

The factors which influence the depth of cyanosis caused by a given quantity of reduced hemoglobin

(a) The state of cutaneous capillaries. When these are dilated more of the dark colored blood will be present in the skin than when they are constricted. In the former instance the cyanosis will of course be more pronounced. Increased carbon dioxide tension in the blood causes capillary dilatation, therefore, when retention of this gas accompanies oxygen want as in obstruction of the trachea, emphysema, venous congestion of superficial regions, etc. the cyanosis is intense. When, on the other hand, the carbon dioxide tension is low the cutaneous vessels are narrowed. In the anoxemia resulting from rapid shallow breathing (p. 426) therefore the cyanosis tends to be of a pale leaden or heliotrope hue, if the peripheral vessels are strongly constricted, as in shock, cyanosis may be inconspicuous.

(b) Pigmentation and thickness of the skin. These factors obviously will modify the depth of the cyanotic color. Cyanosis is more clearly evident in regions where the skin is thin and unpigmented. The yellow discoloration of the skin caused by an excess of bilirubin in the blood (jaundice) tends to modify the cyanotic tint, but since the former stains the skin itself, while the discoloration due to reduced hemoglobin is confined to the capillary vessels, jaundice is likely to be just as intense in regions where the skin is thick as in those where it is

thin. The cyanotic discoloration can be temporarily abolished by pressure upon the skin whereas the icteric staining cannot. Cyanosis does not appear in the conjunctivae but these are deeply colored in jaundice.

OXYGEN THERAPY

Oxygen administration is employed for a greater number of conditions today than it has been in the past. In combating acute arterial anoxia (anoxemia) as may occur in pneumonia, pulmonary edema, or obstruction to breathing, it is of utmost value, as well as in states such as congestive heart failure, or coronary thrombosis when, though the arterial blood contains the usual amount of oxygen, the tissues, owing to impairment of the peripheral circulation, suffer from a deficiency of oxygen.

The suitability of a given case for the administration of oxygen depends upon certain quite definite principles. The chronic anoxia due to anemia, in which the hemoglobin is saturated with oxygen to the normal degree, is treated preferably by measures directed to the disease itself rather than by oxygen therapy. In failure of the peripheral circulation, the inhalation of 100 per cent oxygen will increase the oxygen content of the blood to revive a flagging respiratory center. But the anoxia due to a shunt of blood from the right to the left side of the heart (p 432) or through a similarly completely unaerated portion of the lung will not be much benefited by oxygen inhalations, it is not possible to make the blood supplying healthy and well aerated alveoli absorb any important amount of extra oxygen and so compensate for the shunted blood.

When, however, the diffusion coefficient of oxygen (p 366) through the alveolar membrane is reduced as a result of edema, thickening or a coating of fluid, oxygen administration by raising the pressure of the gas in the alveolar air will increase its rate of diffusion across the pulmonary epithelium. The oxygen saturation of the blood flowing through the damaged pulmonary tissue is increased.

Consequently, in broncho- or lobar pneumonia when such changes are responsible for the anoxemia, in emphysema or in pulmonary edema whether from cardiac or pulmonary disease or from gas poisoning, the success of oxygen administration is often spectacular. Anoxemia increases the permeability of the pulmonary epithelium to fluids

and, consequently encourages edema formation. In other words, a vicious circle is established—edema inducing anoxemia and the latter increasing the edema—which is broken by oxygen administration.

Anoxemia due to rapid shallow breathing is relieved by oxygen treatment since the oxygen tension of poorly ventilated alveoli is raised thereby. The shallow breathing itself is likely to persist since it is primarily due to the local process in the lung acting upon the nerve endings rather than to the anoxemia. Therefore this type of breathing could, no more than the pulmonary lesion itself, be expected to be abolished by oxygen treatment. Nevertheless, since shallow breathing is aggravated by oxygen want it will be ameliorated by oxygen treatment—the vicious circle is broken.

It is often said that when the arterial blood contains the normal quantity of oxygen, namely about 19.5 volumes per cent (saturation 97.5 per cent) as in the stagnant type of anoxia of congestive heart failure, oxygen inhalation cannot be of any great value, since the oxygen saturation can be raised only to 100 per cent, and the total oxygen content increased by 2.2 volumes per cent (an increase of 1.7 volumes in simple solution together with 0.5 volumes combined with hemoglobin). That is, the total oxygen content of a person breathing 100 per cent oxygen will be 21.7 volumes per cent. But this increase of 2.2 volumes per cent (representing an increase in O_2 content of 11 per cent) is as pointed out by Comroe and Driggs of the highest importance in anoxia, for it raises very considerably the pressure at which oxygen is delivered to the tissues. Say, for example that, when a patient is breathing ordinary air, every 100 cc. of blood in passing through the capillaries loses 7 volumes of O_2 (normal about 5 vols), and, therefore, leaves the venous end of the capillary with a content of 12.5 vols per cent, is 63 per cent saturated, and has a pO_2 of around 32 mm Hg. Now when 100 per cent oxygen is breathed the extra 2.2 volumes per cent in the arterial blood is given up on entering the capillaries, which reduces the oxygen saturation only to that existing under ordinary conditions in the arterial blood, namely, 97.5 per cent. But the tissues owing to the slowing of the circulation abstract a further 4.8 volumes per cent to satisfy their need of 7 volumes per cent. The venous blood, therefore, contains 14.7 volumes per cent of oxygen, and is over 73 per cent saturated, instead of about

63 per cent without oxygen administration, the pO_2 is nearly 40 mm Hg. Thus the tissues are supplied with oxygen at a virtually normal head of pressure. In congestive heart failure, therefore, especially if there is generalized edema, the administration of oxygen is often of great benefit.

Methods of administration

Oxygen is given usually in a concentration of from 70 to 100 per cent by means of a nasal or oral-nasal catheter, by a specially designed face-mask, or the patient is placed in a hood-tent or air-tight cabinet in which the O_2 percentage is maintained at the required concentration. Carbon dioxide (5–10 per cent) is sometimes added, especially in CO poisoning, persistent hiccough and postoperative atelectasis, in order to encourage lung expansion.⁴

Effects of the inhalation of 100 per cent oxygen

In normal persons the breathing of 100 per cent oxygen often causes a slight initial reduction of breathing of about 3 per cent. This is attributed to the removal of the tonic action of the chemoreceptors of the carotid and aortic bodies (glomera) which are stimulated evidently by the unsaturation (2.5 per cent) of the blood existing at sea-level. If the administration is continued the respirations are stimulated as a result of afferent impulses set up in the respiratory tract by the inhalation and conveyed to the medullary center. In anoxia, a much more pronounced depression or even complete cessation of breathing may occur. This so-called *oxygen apnea* may be very alarming, especially during anesthesia. But it is a sure indication of the existence of anoxia and the need for oxygen therapy, for it shows that the respiratory center has been profoundly depressed to the direct stimulating action of CO_2 . Before oxygen administration the center had been driven by impulses from the peripheral chemoreceptors, the relief of the anoxia has abolished this source of stimulation. In chronic anoxia, such as may be seen in congestive heart failure and pulmonary emphysema, especially when there is also carbon dioxide retention, the patient may pass into coma when oxygen is administered. The loss of consciousness is the result of withdrawal of the stimulating effect of the anoxia upon the peripheral

⁴ The patient's own breath adds from 1 to 2 per cent to the atmosphere of the tent.



FIG 34 10 The Barach-Davidson oxygen tent.

chemoreceptors and the consequent acute anoxia of the cerebral cortex. The carotid and aortic bodies which had been under stimulation by both anoxia and the high pCO_2 of the arterial blood are, after oxygen is inhaled, dependent upon CO_2 alone for their excitation, while the respiratory center itself has become tolerant, i.e., insensitive to carbon dioxide. The high level of CO_2 in emphysema (sometimes as high as 150 volumes per cent) also exerts a narcotic effect upon the cortical cells.

Another effect of oxygen inhalation is the elimination of nitrogen. A man of average weight (70 kg) breathing oxygen eliminates about 18 cc of nitrogen per minute. The rate of removal of the gas from all tissues is not however the same, it is removed most rapidly from the blood and the most vascular parts of the brain. Pure oxygen is, therefore, administered to divers to prevent decompression sickness, and to pilots before high altitude flights (p. 425).

High concentrations of oxygen, especially under pressures of more than one atmosphere, may exert irritant effects upon the respiratory tract and has certain other deleterious actions. These, which will be merely listed, include substernal distress, reduced vital capacity, bronchitis. At O_2 inhalations at a pressure of more than one atmosphere neurological symptoms may develop.

Helium (atomic weight 4, $\frac{1}{8}$ density that of nitrogen) is lighter than any other gas except hydrogen. Barach has applied this physical fact to reduce the respiratory effort in asthmatic attacks, in those suffering from laryngeal or tracheal obstruction and in certain other types of dyspnea. A gas mixture is used in which helium is substituted for nitrogen (oxygen 21 per cent, helium 79 per cent). For the relief of anoxemia the oxygen percentage may be increased to 60 or 70 per cent.

SECTION IV THE EXCRETION OF URINE

By N B T

CHAPTER 35

URINE FORMATION THE STRUCTURE OF THE KIDNEY, THEORIES OF RENAL FUNCTION, VOLUME AND COMPOSITION OF URINE

THE STRUCTURE OF THE HUMAN KIDNEY

THE NEPHRON This is the functional unit of the kidney. It comprises (1) the *renal (or Malpighian) corpuscle*, and (2) the *renal tubule*, which is divisible into (a) the proximal convoluted tubule, (b) the loop of Henle, and (c) the distal convoluted tubule. There are about 1 million nephrons in each human kidney (figs 35.1 and 35.2).

(1) **THE RENAL CORPUSCLE** consists of what appears to be a twisted skein of capillary channels—the *capillary tuft or glomerulus*—which in development has become invaginated into the upper blind end of the primitive renal tubule. The narrow, funnel-like cavity resulting from this invagination, and which almost completely surrounds the capillary tuft is known as *Bowman's capsule* (fig 35.1). The capsular wall therefore consists of a *visceral* and a *parietal* layer. The former is a delicate membrane of flat cells. It envelopes each capillary loop and blends with the vascular wall. This layer becomes folded upon itself and continuous with the parietal layer at the point where the afferent and efferent vessels (see below) enter and leave the glomerulus. The cells of the *parietal layer* are for the most part squamous in type but become cuboidal near the point where the capsule empties into the tubule. The renal corpuscle measures about 200 microns in diameter. The total surface of the capillary loops of both kidneys is about 1.5 square meters. The renal corpuscles and the convoluted tubules lie mostly in the *cortex* of the kidney.

THE PROXIMAL CONVOLUTED TUBULE is a tortuous tube (about 55 μ in diameter and 14 mm long) lying in close relation to the renal corpuscle, and into which the latter empties. Its walls are composed of a single layer of cuboidal cells which differ from those lining other parts of the renal tubule in possessing delicate striations perpendicular to their free borders—the brush or bristle border. This appearance is suggestive of the striations seen in the epithelial cells lining the small intestine (Cowdry). The basal portions of the cells show a reticulum of protoplasmic threads radially arranged—the so-called *rods*. The length of the proximal tubule when

uncoiled is some 15 mm. Its lumen varies in diameter from 15 to 25 microns in accordance with the quantity of fluid passing through it. The combined area of the 1,000,000 proximal tubules in a human kidney is nearly 1 square meter.

HENLE'S LOOP follows the proximal convoluted tubule and consists of a *descending* and an *ascending* limb. The distal third or so of the loop dips into the renal medulla. The proximal four-fifths of the descending limb, though it follows a nearly straight or a moderately tortuous course, has about the same diameter as that of the proximal convoluted tubule and is lined by similar cells. The distal fifth of the descending limb becomes greatly narrowed and is lined by clear flat cells. The function of this part of Henle's loop is not definitely known (see p. 450). Beyond the narrowed portion the tubule (ascending limb) widens again to its previous diameter. This portion is lined by cuboidal or columnar epithelium and is continued into the distal convoluted tubule. The average length of Henle's loop is about 20 mm.

THE DISTAL CONVOLUTED TUBULE resembles the proximal convoluted tubule and lies coiled in close relation to the renal corpuscle. Its cells, however, have no brush border but resemble those forming the ascending limb of Henle's loop. Its length when uncoiled is about 5 mm. The distal tubule comes into contact over a limited area with the wall of the efferent vessel just before the latter enters the glomerulus. At this point the lining cells of the tubule assume a columnar shape, their nuclei are densely packed. The whole structure has a plaque-like appearance and is known as the *macula densa*, the term first applied to it by Zimmermann. This part of the afferent arteriole also undergoes a conspicuous structural change (p. 441). The distal convoluted tubule drains into the collecting duct system.

The length of a single nephron when straightened out is between 1 and 1½ inches (30–38 mm), and a conservative estimate of the total length of the tubules of both kidneys from Bowman's capsule to the first collecting duct is some 40 miles. The total tubular surface cannot be far from 6 square meters.

The nephrons whose glomeruli lie in the outer two-thirds or so of the cortex, and to which the foregoing description especially applies, differ in certain respects

¹ Some investigators have obtained a figure considerably higher than this.

from those in the deeper third. In the latter region, called *juxtamedullary* by Heggie, the glomeruli are somewhat larger and the tubules, which lie almost entirely in the medulla, much longer. Moreover the loops of Henle are elongated as compared with those of the more superficially placed nephrons, and their thin segments constitute a much greater proportion of the total tubular length. In the more superficial nephrons this part of Henle's loop, as mentioned above, is only about $\frac{1}{4}$ of the total length of the tubule, whereas in the juxtamedullary nephrons the length of this part is nearly one-half of the total and is continued into the ascending limb of the loop. The blood supplies of the two types of renal unit also show distinctive features (p. 442). The juxtamedullary nephrons constitute about 15 per cent of the total number of the nephrons in the kidney.

A SYSTEM OF COLLECTING TUBULES conveys the urine to the kidney pelvis. These tubules have generally been considered to possess no function beyond acting as mere conduits for the urine, and have therefore not been included as part of the nephron. But there is now evidence that they absorb water and, thus, have some concentrating power. The smallest of the collecting tubules (the *initial* or *connecting tubules*) receive urine from the distal convoluted tubules. Several connecting tubules from neighboring nephrons join to form a larger channel. Through a succession of such unions, relatively large short tubes—the *papillary ducts* (or *ducts of Bellini*) are finally formed which open into the renal pelvis at the apices of the papillae (fig. 35.2).

BLOOD SUPPLY The *renal artery* upon entering the hilus of the kidney breaks up into numerous branches—the *interlobar arteries*—which pass outward between the renal pyramids to the junction of the cortex with the medulla. Here they turn to follow a more horizontal course and form arterial arches across the bases of the pyramids. From these vessels—called *arcuate arteries*—arise the *interlobular arteries* which run outwards through the cortex for variable distances. From this point onwards the circulations of the superficial and the juxtamedullary nephrons are separate and different.

The *superficial glomeruli* receive twigs from the interlobular arteries at intervals along the latter's outward course. Each of these offshoots with few exceptions enters a renal corpuscle and constitutes the *afferent vessel* of the glomerulus. The afferent vessel throughout the greater part of its course has the features of an arteriole, it contains muscle fiber and has a diameter of about 50 microns. Upon entering the renal corpuscle and breaking up into a limited number of primary and secondary branches it gives rise to a tangle of some 50 capillary loops which constitute the glomerular tuft. *There is no anastomosis between any of these loops, and each is enveloped in a prolongation of the visceral layer of Bowman's capsule*, much as the small intestine is covered by its serous coat, or as Vimtrup expresses it, "as a finger is covered by a glove." Nothing therefore separates the blood in each capillary from the cavity

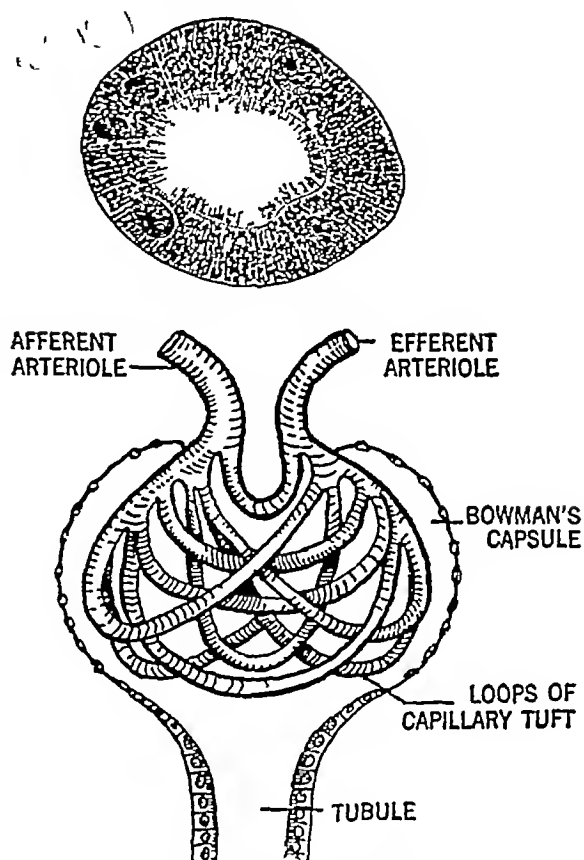


FIG. 35.1 Upper drawing, cross-section through the convoluted tubule of a dog's kidney, showing brush border (From Maximow and Bloom). Lower drawing, diagram of Malpighian corpuscle. Note that capillaries pass from the afferent to the efferent arteriole in separate loops. There are many more capillary loops than are shown here.

of Bowman's capsule except two delicate membranes which together have a thickness of about 1 micron. The capillary loops after a course of some 0.5 mm converge to form the *efferent vessel*, which leaves the renal corpuscle close to the point of entrance of the afferent vessel. The diameter of the efferent vessel is only about one-half that of the afferent, its sectional area is therefore only about one-quarter that of the latter (see figs. 35.3 and 35.4).

The wall of the afferent vessel just before it enters the glomerulus loses its elastic membrane, the endothelium becomes discontinuous and the muscle fibers are overlaid and in part replaced by a cushion of myoepithelial cells (the *polkissen* or "*pole-cushion*" of Zimmerman). We have seen (p. 440) that the distal tubule comes into contact with this part of the afferent vessel. Cells possibly of a neural or secretory character are found, as well, in the angle between the afferent and efferent arterioles. These, together with the myoepithelial cells in the wall of the afferent vessel constitute what has been termed the "*juxta-glomerular apparatus*." This structure, which also contains numerous

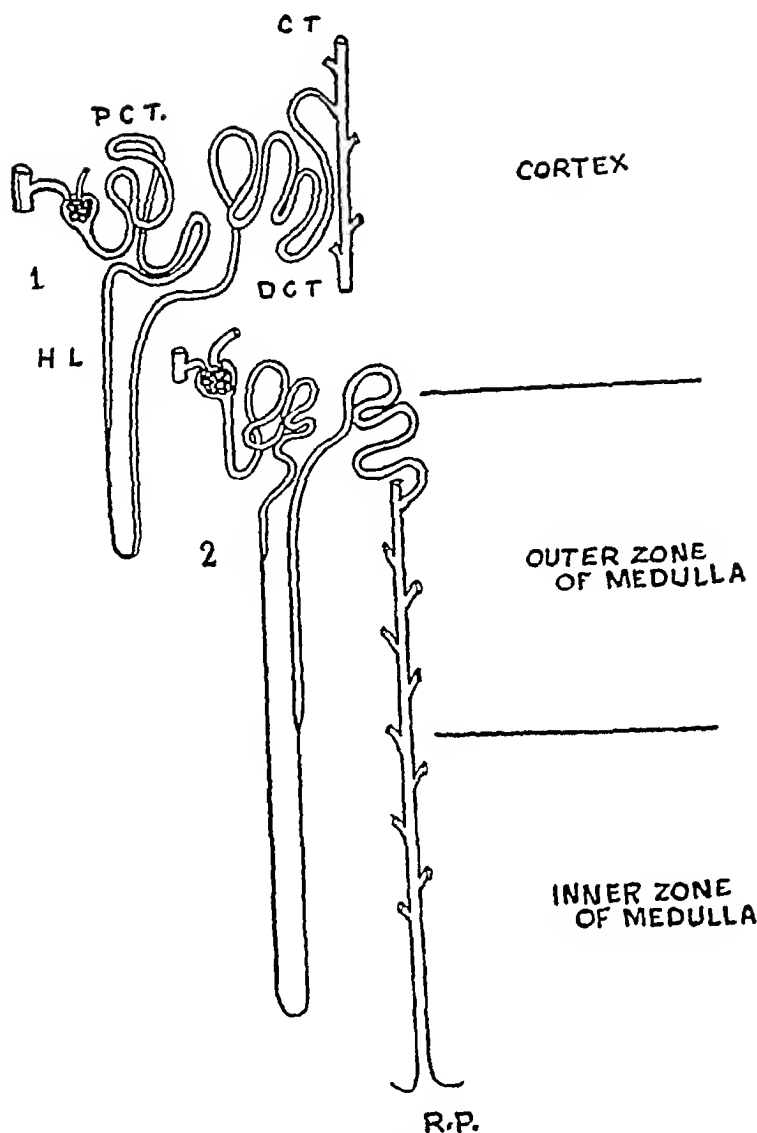


FIG 35 2 Diagram of a cortical (1) and of a juxtamedullary (2) nephron P.C.T., proximal convoluted tubule, D.C.T., distal convoluted tubule, C.T., collecting tubule, H.L., Henle's loop, R.P., renal pelvis

nerve fibers, has been thought by some to serve an endocrine function, namely, the regulation of the glomerular blood flow by the liberation of renin (p 157). The macula densa is possibly associated functionally with the juxtaglomerular apparatus (p 441).

In the *juxtamedullary region*, the afferent arteriole arises from an interlobular artery near its origin from the arcuate artery, or directly from the latter vessel itself. The efferent arterioles break up into a number of straight, relatively large calibered vessels—the *arteriolar rectae* or *vasa recta*—which course through the medulla.*

* Some of the afferent arterioles have been described as giving off a communicating vessel (Isaacs-Ludwig

There is much less disparity between the diameter of afferent and efferent vessels of the glomeruli of this

arteriole) which by passes the glomerulus and connects with the capillary plexus surrounding the tubules. Occasionally an interlobular artery also gives off a branch which joins the capillary plexus directly. The number of these extra glomerular vessels and consequently their functional importance is a matter of dispute. They are said to increase in size and number in old age, as well as in chronic kidney disease and then probably play a more important rôle than in health. Trueta and his colleagues believe that they are always pathological and are formed simply by the continuation of the afferent into the efferent arteriole, the glomerular capillaries, except for a single expanded channel, having degenerated and disappeared.

region than is seen in the corresponding arterioles in the outer part of the cortex. The vasa recta, when injected or filled with blood, appear as dark, straight lines converging toward the apices of the pyramids of the medulla. For the most they enter venous channels without passing through a peritubular capillary network, and many after a short course loop sharply backwards (proximally) to reach their destinations. The walls of these vessels are capillary in structure, that is, they consist of a single layer of endothelial cells.

Under ordinary circumstances the renal blood flow passes almost entirely, it is believed, through the more superficially placed glomeruli. But, as Trueta and his associates have found in animals, the vessels of the juxtamedullary nephrons provide a "by-pass" through which blood may be diverted, and thus leave the cortex partially, or completely, deprived of blood. Such a diversion has been shown to occur in tourniquet shock (ch 27), or as a result of prolonged afferent nerve (sciatic) stimulation, and in traumatic anurias (e.g., the crush syndrome) in man. The time taken for the passage of blood from renal artery to renal vein is then much reduced, and bright stream-lines appear in the venous blood (fig 36 1, p 479).

The kidneys of frogs, fish and snakes have a double blood supply. The renal artery supplies the glomeruli, but the capillary network around the tubules is derived from the renal portal vein.

THE VENOUS RETURN The blood having traversed the capillary vessels surrounding the convoluted tubules is collected into a venous plexus in the cortex. The blood from this plexus passes successively through interlobular, arcuate and interlobar veins which accompany the corresponding arteries. The interlobar veins become confluent near the hilus of the kidney to form the renal vein.

There are certain anatomical features of the renal circulation of special physiological interest which should be emphasized.

(1) Nearly all the blood which enters the kidney passes through the glomerular tufts.

(2) The tubules of the normal kidney are therefore supplied largely by blood which has first passed through the glomeruli.

(3) The renal artery breaks up quite suddenly into short branches. This fact, to which attention was first drawn by Bowman, ensures that blood is delivered to the glomeruli under a high head of pressure. It should also be remembered that the renal artery is short, arises directly from the aorta, and is very large proportionately to the mass of tissue which it supplies.

(4) There is a striking disparity between the diameters of the afferent and efferent vessels.

(5) There are two separate circulations in the kidney, a *greater*, which ordinarily carries 85 per

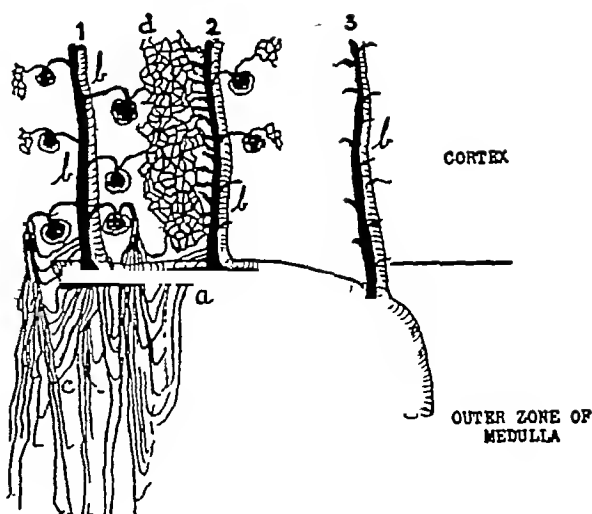


FIG 35 3 Diagram of the renal circulation, *a*, arcuate vessels, *b*, interlobular vessels (1, 2 and 3), arteries black, veins shaded, *c*, vasa recta, *d*, peritubular plexus of vessels. In 2 and 3 peritubular plexus and all but two glomeruli are omitted.

cent of the blood and supplies the more superficially placed glomeruli, and a *lesser*, which distributes blood to the juxtamedullary nephrons and may serve upon occasion as a "by-pass".

The renal nerves

The kidney receives a sympathetic and a parasympathetic innervation.

(1) The **SYMPATHETIC FIBERS** arise from the sixth thoracic segment to the third lumbar inclusive. Functionally the fibers are of two types, *vasoconstrictor* and *afferent*. They are conveyed to the kidney in the *greater*, *lesser* and *least splanchnic nerves*. The greater and lesser splanchnic fibers synapse in the semilunar ganglion. The postganglionic fibers upon leaving the ganglion enter into the formation of the *renal plexus*, in which fibers of the vagus are also intermingled. From this plexus, which surrounds the renal vessels, non-medulated fibers may be traced to the arterioles and to the glomerular and tubular capillaries. Nerve filaments have even been described between the cells of the tubules and of Bowman's capsule. The presence of ganglion cells within the substance of the kidney is a disputed question. The fibers of the least or lowest splanchnic nerve connect with cells within the renal ganglion situated in the hilus of the kidney, postganglionic fibers enter the renal substance.

(2) The **PARASYMPATHETIC FIBERS** are derived from the vagus. Their functions are unknown. The evidence is against either this nerve or the sympathetic supplying true secretory fibers to the renal cells. The effects upon urine formation caused by section or by the stimulation of any of the renal nerves are believed to be entirely the result of the vascular change which such section or stimulation has brought about. Carrel and Guthrie

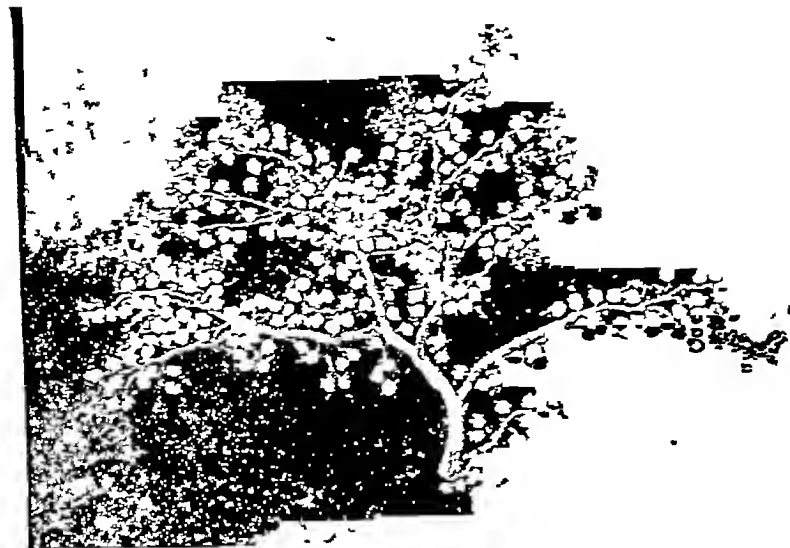


FIG 354. Cast of vessels and glomeruli of human kidney prepared by injecting neoprene and subsequently corroding with acid. (courtesy of Professor G. L. Duff)

and Quinby excised the kidney and transplanted it into another position. Though completely denervated by this procedure, the kidney excreted urine of the usual composition.

The hypothalamus and the orbitofrontal part (area 13) of the cerebral cortex is capable of affecting the renal circulation. Stimulation of either of these regions causes a shunt of blood from the cortical to the juxtamedullary glomeruli (Cort)

THEORIES OF RENAL FUNCTION

Borrmann (1842) discovered the fundamental fact that the capsules surrounding the glomerular tufts are simply the expanded extremities of the renal tubules. Struck by the design of the glomerulus, which seemed to fit it so admirably to the purposes of a filter, he believed that this structure filtered water but that the solids of the urine were *secreted* by the cells of the tubules. Heidenhain maintained that the formation of urine was entirely a secretory process. Water and salts were the products of the glomerular membrane, while the more characteristic urinary constituents, urea, uric acid, etc., and under certain conditions water as well, were secreted by the tubular epithelium (see p. 451). He and his followers have endeavored by almost countless experiments—largely involving the injection of dyes and the demonstration that these appeared in high concentration in the tubule cells—to prove the secretory function of the tubules. The theory of Ludwig was based upon

purely physical conceptions. He believed that the glomerulus acted as a filter to remove from the plasma, water and all substances other than the proteins, and that the filtrate was concentrated in the tubular lumen through the return of the greater part of the water to the blood, by means of a simple process of diffusion.

Cushny's "modern theory"

It is quite obvious that Ludwig's theory cannot be correct in its entirety, if filtration be granted, then the several urinary solids should not only be the same as those of plasma (less the proteins) but, if water alone underwent reabsorption, they should all be concentrated in the urine to precisely the same degree. We know that this is not so. Urea, for instance, is concentrated some 60 times, creatinine 70 times or more, whereas chloride may be concentrated only twice, and sugar appears in the urine in traces, if at all (see table 34). Either simple filtration must be deleted from Ludwig's theory and selective secretion substituted, or the conception of the reabsorption of water alone must be replaced by one postulating the reabsorption of water and certain solids in different relative proportions. Cushny's theory retains simple filtration by the glomerulus but postulates such a *selective re-absorption* by the tubules. According to this view the reabsorbed fluid is of constant composition under all conditions of health and resembles Locke's solution i.e., one containing sodium, cal-

TABLE 34

Showing composition of plasma, glomerular filtrate, reabsorbed fluid and urine (modified from Cushing)

	90 LITERS PLASMA CONTAIN		83 LITERS FILTRATE CONTAIN	82 LITERS REABSORBED FLUID CONTAIN		1 LITER URINE CONTAINS	
	Per cent	Total		Per cent	Total	Per cent	Total
Water	92	83 liters	83 liters		82 liters	95	950 cc
Colloids	7.5	6750 gr					
Glucose	0.1	90 gr	90 gr	0.11	90 gr		
Sodium	0.3	270 gr	270 gr	0.32	266.5 gr	0.35	3.5 gr
Chloride	0.37	333 gr	333 gr	0.4	327 gr	0.6	6.0 gr
Urea	0.03	27 gr	27 gr	0.008	7 gr	2.0	20.0 gr
Uric acid	0.004	3.6 gr	3.6 gr	0.003	3.1 gr	0.05	0.5 gr
Potassium	0.02	18.0 gr	18.0 gr	0.02	16.5 gr	0.15	1.5 gr
Phosphate	0.009	8.1 gr	8.1 gr	0.0008	6.6 gr	0.15	1.5 gr
Sulphate	0.002	1.8 gr	1.8 gr			0.18	1.8 gr
Creatinine	0.001	0.7 gr	0.7 gr			0.07	0.7 gr

cium, potassium, magnesium, chlorine and sugar, and in addition, variable quantities of urea, uric acid, and phosphate. A selective process of this nature cannot be explained upon the basis of known physical laws, but entails "vital" activity, i.e., the performance of work by the tubular epithelium. Filtration, on the other hand, is "due to a blind physical force." The glomerular membrane plays a passive rôle, performing no work (see table 34). These two processes, glomerular filtration and tubular re-absorption will be considered in detail.

Though the main tenets of the Ludwig-Cushing theory, namely filtration by the glomerulus and reabsorption of water, glucose, essential salts and certain other solids by the tubules, are firmly established, later research has added greatly to our knowledge of renal function, and has required many emendations and additions. The reabsorbed fluid, for example, is not of constant composition, but shows considerable variation.

EVIDENCE FOR GLOMERULAR FILTRATION

(1) **STRUCTURE** The peculiar features of the renal circulation—the great number of capillary loops in the glomerular tuft, the disparity between the calibres of its afferent and efferent vessels, the origin of the renal artery directly from the aorta, and its abrupt division into branches (p. 441) indicate that the development of pressure within the tuft is an important requirement for urine formation, and consequently strongly suggest that this part of the renal unit acts simply as a filter. On the other hand, a close examination of the

glomerulus reveals nothing which might suggest its performing a secretory function.

(2) **THE FILTRATION PRESSURE** In order to drive fluid through a permeable membrane, there must be a higher pressure on one side. Variations in this pressure will alter the volume but not the composition of the filtrate. If urine formation involves filtration, alterations in the glomerular pressure should cause corresponding changes in urine volume. The glomerular pressure, however, depends not only upon the difference between the blood pressure within the capillary loops and the pressure in Bowman's capsule, but also upon the osmotic pressure of the colloids (proteins) of the plasma. This colloid pressure, which amounts to between 25 and 30 mm Hg, opposes the pressure of the blood within the glomerular tuft. In other words, such an osmotic pressure requires that, in order for filtration to occur, the blood pressure shall be 25 or 30 mm higher than it would need to be were the plasma protein-free. The effective filtration pressure, or actual driving force, may be expressed thus—

$$P_b^{(75)} - P_o^{(30)} - P_c^{(5)} = P_f^{(40)}$$

Where P_b = blood pressure, P_o = osmotic pressure of proteins, P_c = pressure in Bowman's capsule and P_f = effective filtration pressure (see fig. 35.5).

Hayman measured the pressure in the afferent arterioles and the capillaries of the frog's kidney simultaneously with the aortic pressure. The method employed in measuring the pressure in the renal vessels was one based upon the general principle employed in human blood pressure esti-

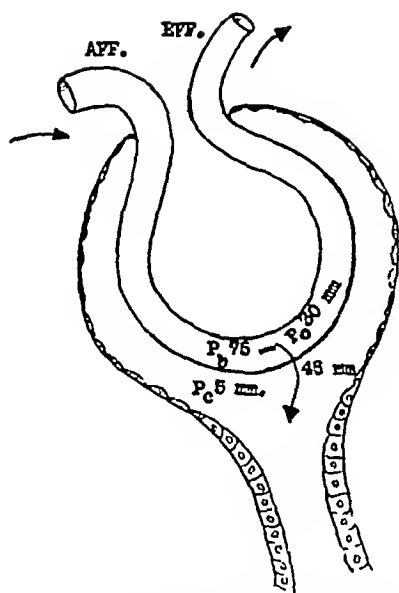


FIG. 33.5 Diagram illustrating the relation of blood pressure (P_b), osmotic pressure (P_o) and capsular pressure (P_c) upon the formation of urine. Aff = afferent arteriole, Eff = efferent arteriole

mations. A fine pipette was inserted into Bowman's capsule and the proximal convoluted tubule occluded by compression with a glass rod. Fluid was then run into the capsule under measured pressures. The pressure at which blood was observed to enter the vessels of the tuft, during systole only, was taken as the index of the systolic pressure in the afferent vessel. As the pressure was gradually lowered below this level the flow in the capillaries became continuous. The pressure at which this occurred was taken to correspond to the pressure (diastolic) in the capillary loops. The following mean values were obtained:

Systolic pressure in aorta	37.4 cm H_2O
Systolic pressure in afferent vessel	31.6 cm H_2O
Diastolic pressure in capillaries	20.2 cm H_2O , or 54 per cent of the systolic aortic pressure

Capillary pressures as high as 70 per cent of the aortic pressure were observed.

Numerous experiments of various types have been carried out to investigate the effects of one or other of the three factors mentioned above (P_b , P_o and P_c) upon the formation of urine. In general the results accord with a filtration process.

(a) *Blood pressure* Raising the general blood pressure causes a corresponding change in the urinary flow, provided that the agent which raises the systemic pressure does not cause a reverse change in the blood pressure within the kidney itself. Stimulation of the splanchnic nerves, for example, with the renal nerves intact is likely to result in a reduction in urine volume from the kidney of that side, since its vessels share in the general vasoconstriction and the glomerular pressure is reduced. If the renal nerves are first cut, however, the general vasoconstriction and the renal vasodilatation (removal of constrictor tone) caused a higher pressure within the vessels of the kidney and a profuse flow of urine. Division of the cord in the thoracic region, by reducing the general blood pressure through the removal of the vasoconstrictor tone, reduces the urinary flow. The fall in blood pressure following hemorrhage acts similarly.

The rate of the renal blood flow as well as the pressure within the glomerulus will be altered by the foregoing measures, and the altered blood flow, rather than the change in blood pressure, might be held responsible for the greater urine production. It was shown, however, by Richards and his associates that variations in glomerular pressure in itself caused parallel changes in urinary flow when special means were employed to keep the renal blood flow constant. It was also shown by these observers that in certain dosage adrenaline causes a relatively greater constriction of the efferent than of the afferent vessel. A rise in glomerular pressure and an increased flow of urine resulted.

(b) *Osmotic pressure of proteins* As already stated, this amounts to between 25 and 30 mm Hg. The blood pressure in the capillaries of the tuft must therefore exceed the osmotic pressure in order that filtration shall occur (see also p. 31). The blood pressure in the capillary loop is around 60 per cent of the systolic aortic pressure, or from 70 to 80 mm Hg. So then, over and above the sum of osmotic and intracapsular pressures, there exists a capillary pressure of from 35 to 50 mm Hg which serves to drive fluid across the glomerular membrane. Reduction of the blood pressure by this amount (from 35 to 50 mm Hg) suppresses urine formation.

The intravenous injection of isotonic salt solution causes profuse diuresis, resulting from, in part at least, the dilution of the plasma proteins, saline whose osmotic pressure has been raised by the addition of gelatin or gum acacia causes a diuresis only about one-third as great (fig. 35.6). Dilution of the plasma colloids permits the blood pressure to be reduced to a much lower level (under 20 mm) before urine formation ceases. It is also well known that the urine volume is less in the standing position than in recumbency, this is attributed in part to the concentration of plasma protein which occurs in the former position (p. 26).

(c) *Intracapsular pressure* Under ordinary circum-

stances this is low, probably not more than from 5 to 10 mm Hg. It is largely due to the pressure of the surrounding renal substance (intrarenal pressure), for the kidney, it will be recalled, is enclosed in an inelastic fibrous capsule. When a manometer is tied into the ureter, the pressure in the tubules and capsule gradually rises. Urine formation ceases when the instrument registers a pressure equal to that by which the glomerular blood pressure exceeds the colloid osmotic pressure, i.e., when $P_o(30 \text{ mm}) + P_c(45 \text{ mm}) = P_b(75 \text{ mm})$. Dilution of the proteins will permit the ureter pressure to rise higher before urine formation ceases. The effect of raising the ureter pressure in suppressing urine formation is, of course, strong evidence for filtration, a true secretory process is not dependent upon a difference between the pressure in the vessels and that in the secretory ducts. The salivary gland, for example, continues to secrete though the duct pressure greatly exceeds the capillary blood pressure.

(3) COMPOSITION OF THE CAPSULAR FLUID Wearn and Richards observed the glomeruli of the frog's kidney beneath the microscope and drew off fluid from Bowman's capsule by means of a fine pipette (fig 35 7). *The fluid had the composition of protein-free plasma.* It was alkaline in reaction and contained urea, sodium chloride and glucose in virtually the same concentrations as in plasma. Dyes injected into the animal appeared in the capsular fluid in the same concentration as in protein-free plasma. The molecular concentration and electrical conductivity of the capsular fluid and of de-proteinized plasma were also the same, the capsular fluid is therefore a filtrate. Walker and his associates

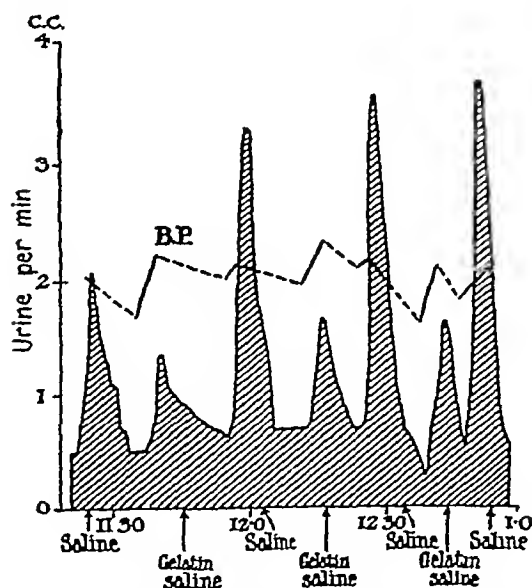


FIG 35 6 Illustrating the effect of the intravenous injection of saline and of a gelatine saline solution, respectively, upon the flow of urine. BP = blood pressure. The figures along the abscissa indicate the time in half-hours (After Knowlton.)

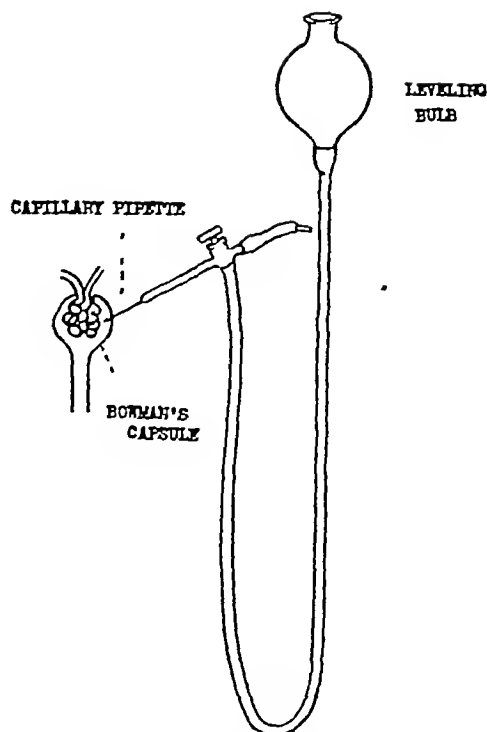


FIG 35 7 Diagram showing Wearn and Richards method of obtaining a sample of fluid from Bowman's capsule. A fine glass rod (not shown in figure) is used to compress the proximal tubule just below the capsule and so to prevent fluid from being drawn from the tubule. When the glass bulb and tubing containing mercury is lowered, capsular fluid alone is withdrawn.

have succeeded in recovering samples of fluid for analysis from the capsule and tubules of the mammalian kidney (guinea-pig and rat). The observations of Wearn and Richards upon the amphibian nephron have in general been confirmed.

(4) THE PLASMA PROTEINS when they pass into the urine do so in amounts which are inversely related to their molecular weights, showing that secretion plays no part, but that molecules escape through "pores" in the glomerular membrane. In kidney disease, serum albumin, which has a molecular weight of around 70,000, appears in the urine, as a rule, in greater amounts than does serum globulin (molecular weight about 165,000).³ Fibrinogen with a still larger molecule rarely escapes across the glomerular barrier. Hemoglobin, which has a molecular weight of a little over 68,000, when free in the plasma passes in the urine fairly rapidly,⁴ whereas hemocyanin, which has a huge

³ Osmotic pressure determination.

⁴ Monke and Yuile estimate that only about 3 per cent of the "pores" of the glomerular membrane are sufficiently large to allow the passage of the hemoglobin molecule. It is possible, however, that the relatively free excretion of hemoglobin is due to the fact that a proportion of the molecules are of smaller size than 68,000. These investigators find that part of the hemoglobin filtered through the glomerulus is reabsorbed from the tubules by phagocytosis.

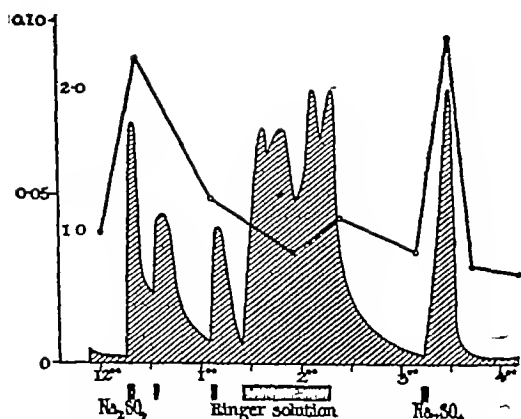


FIG 35.8 Graph of oxygen consumption by the kidney during diuresis following the intravenous injection of Ringer's solution and sodium sulphate. The black line represents the oxygen in cc. taken up by each gram of kidney per minute. The shaded areas denote the volume of urine secreted each minute. Time in half hours. A marked rise in oxygen consumption results from sodium sulphate but not from Ringer's solution (After Barcroft and Strauh)

molecule (mol. wt. 5,000,000) is not excreted. Inulin (mol. wt. 5200) and certain foreign proteins, such as egg white and gelatin with molecular weights between 35,000 and 45,000, and the abnormal Bence Jones protein (p 474) with a molecule of about the same size as the former two proteins are freely excreted. The "pores" of the glomerular membrane appear, therefore, to be of such a size as to allow the passage of molecules of a weight less than about 70,000 (see Bayliss and associates and Bott and Richards). Hemoglobin and serum albumin are, therefore, near the border line, the former appearing in the urine only when its concentration in the plasma rises above 100 mg per 100 cc. Serum albumin may escape into the urine as a result, apparently, of even a slight change in the permeability of the glomerular membrane as in functional albuminuria (p 474).

(5) **OXYGEN CONSUMPTION** Certain measures, such as the injection of saline or the elevation of blood pressure, cause profuse diuresis but little or no increase in oxygen consumption (fig 35.8). This is further evidence that physical processes are concerned, rather than secretion, the latter, of course, would entail oxidative processes and the expenditure of energy. Van Slyke and associates found that the oxygen consumption of the dog's kidney varied widely—from 0.04 to 0.30 cc. per gram of renal tissue per minute—but no relation between the oxygen consumption and the quantity of water or urea excreted was observed.*

* The experiments were made upon unanesthetized animals, one kidney was removed, the other, upon which the observations were made, was transplanted to a subcutaneous position.

It has been calculated that only about 1 per cent of the oxygen consumed by the kidney is used in providing energy for its excretory function (reabsorption). The remainder is used in non excretory processes. It may also be mentioned incidentally here that the blood in the renal vein is usually more than 85 per cent saturated with oxygen, indicating that the oxygen tension is maintained at a higher level in the renal tissue than in most other tissues.

The evidence in the foregoing paragraphs is conclusive for the view that the mammalian glomerulus acts simply as a filter.

Smith and his associates found by means of the inulin clearance method (p 471) that the quantity of filtrate formed by the human kidney was around 125 cc per minute. At this rate nearly 200 liters of fluid would be filtered in 24 hours. Of this, about 99 per cent is reabsorbed by the tubules.

TUBULAR REABSORPTION

Since glomerular filtration is an established fact, some measure, at least, of reabsorption by the tubules must be admitted. This is self-evident, inasmuch as bladder urine contains less water, no sugar and usually less chloride than the glomerular filtrate does. Wearn and Richards found chloride and glucose absent from frogs' urine, yet these substances were present in the capsular fluid. Also, the volume of the urine was only about $\frac{1}{10}$ to $\frac{1}{20}$ of the calculated volume of the filtrate. Water, chloride and sugar quite evidently were reabsorbed from the filtrate during its passage along the tubules. Those substances, such as glucose, amino-acids, vitamin C, chlorine, sodium, potassium and calcium, which are absorbed with water to a relatively large extent and consequently are absent from the urine, or in low concentration, are called *high threshold substances*. High threshold substances, however, appear in the urine in greater amounts if their concentrations in the plasma rise beyond a certain optimal threshold. The glycosuria in diabetes and the hypercalcuria of parathyroid excess (p 858) are examples. *Low threshold substances* are those which are highly concentrated in the urine, since they are reabsorbed in relatively small amounts or not at all. *Urea, uric acid, phosphates and sulphates* are reabsorbed in variable amounts. *Creatinine* is concentrated to a greater extent than any other urinary constituent since it is not normally reabsorbed (see table 34), though in certain pathological states of the kidney, creatinine may undergo passive reabsorption. This

substance is also normally secreted by the cells of the tubules

The main constituents of the glomerular filtrate are grouped by Rehberg into three classes according to the manner in which they behave towards the reabsorption process (a) Substances which are *actively reabsorbed*, namely, those already referred to on page 448 as high threshold substances (glucose, Na, K, Ca, Mg and Cl as well as P and S) These substances are conserved to the body and are usually in lower concentration in the urine and in higher concentration in the reabsorbed fluid than in the plasma (b) Substances which pass back through the tubular epithelium by a simple process of *diffusion* These are the substances already referred to as low threshold substances, namely, urea, uric acid, endogenous sulphate and phosphates They are *passively*, never actively, reabsorbed It has been estimated that in health from 40 to 50 per cent of the filtered urea is returned in this way to the blood These materials are never in lower concentration in the urine nor in higher concentration in the reabsorbed fluid than in the plasma (c) *Non-threshold substances*, which suffer neither active reabsorption nor back diffusion They are therefore absent from the reabsorbed fluid

If reabsorption takes place across tubular membrane, rapid flow of fluid should reduce both active reabsorption as well as the passive process of back diffusion A slower tubular flow should have the opposite effect, such a relationship between the composition of the urine and the time that the filtrate is in contact with cells of the tubules was shown by Cushny When the ureter of one kidney was partially obstructed and the flow along the tubules thereby delayed, the urine from the obstructed side was more concentrated but contained less total chloride than that from the opposite kidney These changes can be accounted for only by a greater reabsorption of salt and water Diuretics, by increasing the flow, caused the production of a more dilute urine but increased the total urinary chloride and urea The more rapid flow along the tubules reduces the time during which the absorption of water and reabsorbable solids can occur Evidence for reabsorption is also provided by the use of dyes Bieter and Hirschfelder injected dye into frogs and observed the kidneys by Richards' method (p 447) The glomerular fluid was stained within 20 minutes Concentration (reabsorption of water) occurred progressively along the tubule, the color being deepest in the distal convoluted tubule

Some other dyes, e.g., trypan blue, are taken up by the cells of the proximal convoluted tubules which become deeply stained After the tubules have been poisoned by mercury or cyanide, concentration does not occur Poisoning of the tubule, however, though it abolishes the specific reabsorptive process permits water and other substances to pass back into the blood by a simple process of diffusion (see passive reabsorption). This latter process has been clearly demonstrated by Richards, who found that complete suppression of urine formation (anuria) may follow the administration of a tubular poison though filtration continues at the normal rate

The proximal convoluted tubules reabsorb between 80 and 90 per cent of the filtered water,* about the same percentage of the filtered sodium and chloride, all the glucose, and most of the other essentially valuable materials, including calcium, potassium, magnesium and vitamin C, potassium is also secreted by the renal cells, probably of the distal tubules, about 85 per cent of the filtered phosphate is reabsorbed, part in the proximal and part in the distal tubule, sulfate is reabsorbed only in minute amounts, less than 0.5 per cent Nearly all the sodium which is not reabsorbed in the proximal tubule is returned to the blood further along the tubule, only about 0.3 per cent appears in the urine The filtrate in the proximal tubule is isotonic, for solids and water are reabsorbed in the same proportion as they exist in plasma In more distal portions of the nephron—the exact site is unknown (see p 450)—the filtrate loses a smaller proportion of solids Here, the remainder of the water of the filtrate, between 10 and 20 per cent, is taken up from the tubular lumen, thus concentrating the solids of the fluid which is now hypertonic and comes to have the composition characteristic of urine The proximal tubules reabsorb a constant proportion of the filtered water, for this reason it is called the *obligatory moiety* The water reabsorbed more distally (averaging 15 per cent) varies in volume and depends upon the activity of the cells of the tubules as influenced by the antidiuretic hormone of the hypophysis and the adrenal cortical principles This fraction of the reabsorbed water is, therefore, termed *facultative*

The urine undergoes acidification in the distal

* The reabsorption of water from the proximal tubule is generally considered to be a *passive* process (that is, one which does not involve the expenditure of energy) It is secondary to the active reabsorption of glucose and salts, and governed by osmotic differences between tubular and peritubular fluids

tubule. Hemoglobin, when injected into the circulation, is deposited as the insoluble acid hematin when acted upon by the acid urine, this has been observed to take place in the distal tubule of the frog's kidney⁷, where acidification of the urine occurs

In its transfer across the tubular membrane, phosphorylation of the glucose⁸ molecule appears to be a necessary preliminary step. The change is brought about by a specific enzyme, *kidney phosphorylase*. Dephosphorylation of the hexose phosphate thus formed is effected by a second enzyme, "*alkaline*" *phosphatase*, present in the cells of the proximal-convoluted tubules

The glomerular filtrate contains small amounts of protein—from 0.5 to 5 mg per 100 cc. Part of this is reabsorbed from the proximal convoluted tubule. The remainder is excreted but normally its amount in the urine is so small—less than 35 mg daily (Addis)—that specially delicate methods are required for its detection. When the passage of protein through the glomerulus is considerable, as in fevers, anemia and certain other states, the reabsorbed protein can be seen as fine globules which give a cloudy appearance to the cells—the so-called "*cloudy swelling*" of pathologists.

Glomerulo tubular balance

The glomeruli as well as the tubules vary considerably in size, and presumably in their capacities for filtration or reabsorption, respectively. But it appears from studies of glucose Tm (p 471), and of glomerular filtration rates, that in the structure of the kidney a nice balance has been set between these two parts of the nephron. Glomeruli of large capacity filter into tubules of similar capacity, and glomeruli of low capacity into low capacity tubules. If there were many nephrons in which this did not obtain, and small glomeruli filtered into tubules of

high reabsorptive capacity then, in order to saturate the entire kidney with glucose (Tm) it would require an extremely high level of plasma glucose, but conversely the presence of many nephrons in which glomeruli of large filtering capacity were associated with small capacity tubules, definite glycosuria would occur at quite low plasma levels of glucose. Such a range in plasma levels between the appearance of glycosuria and complete saturation of the kidney (Tm) is not found in the normal kidney. In renal disease, however, as a result of uneven distribution of the pathological changes in glomeruli and tubules, the normal glomerulo-tubular balance may be upset.

The relation of hormones to the reabsorption of water and electrolytes

The reabsorption of sodium, chloride, potassium and water is influenced by the adrenal cortex and the posterior lobe of the hypophysis (neurohypophysis). The latter exhibits an inhibitory action upon the reabsorption of Na and Cl, but increases the absorption of water. That is, the excretion of salt is increased, while that of water is reduced. In diabetes insipidus, in which the antidiuretic hormone (ADH) of the neurohypophysis is deficient, the excretion of a large volume of urine and a reduced excretion of salt occurs. The injection of a posterior lobe extract (pituitrin) restores the normal urine volume and salt excretion.

The urine volume is varied through the action of ADH on the facultative moiety of the reabsorbed water. We have seen that most of the salt also is reabsorbed from the proximal convoluted tubule, it is thought that the small proportion reabsorbed more distally is alone under hormone control.

Though, as stated above, the exact site of the reabsorption of facultative water is unknown, it is highly probable from indirect or circumstantial evidence that the antidiuretic action is exerted upon the thin segment of Henle's loop. Hypertonic urine is formed by birds and mammals only, and only in these classes does this segment of Henle's loop exist. Moreover, though the antidiuretic principle is not ineffective in other ways upon water metabolism (see Brunn reaction p 794), of those forms, e.g., the frog, in which the thin segment of the loop is absent, it exerts no influence upon the reabsorption of water by their kidneys⁹.

⁹ In the new born infant the urine is hypotonic, or isotonic which may be related to the undeveloped state of Henle's loop which is said to exist at birth. Also, in the young infant the antidiuretic response to

⁷ This fact has a practical bearing when large amounts of hemoglobin are liberated, as may occur after the transfusion of incompatible blood (p 44).

⁸ Phloridzin acts selectively upon the proximal tubule preventing the reabsorption of glucose. This action is the cause of the "diabetes" following the administration of the drug. Phloridzin exerts this action most probably by inhibiting phosphorylase (Kalcar, Kritzler and Gutman), but has no effect upon the "alkaline" phosphatase. Lundsgaard suggested that phloridzin inhibited phosphorylation and dephosphorylation by phosphatase, but several observations have since been recorded which are opposed to this idea. Hudson and Walker, for example, found that iodoacetate which prevents the esterification of glucose by phosphatase did not cause glycosuria, nor is the phosphatase activity of the kidney abolished by phloridzin.

The *adrenal cortex* is essential for the reabsorption of normal quantities of sodium and chloride. In adrenalectomized animals, or in adrenal insufficiency in man (Addison's disease, ch 59) loss of salt is an outstanding feature, the reabsorption of water is also reduced, but the reabsorption of potassium is increased (or its secretion by the tubules reduced). Thus salt and water excretion are increased, while more potassium than normally is retained in the body.¹⁰ These defects are corrected by the administration of cortical steroids. The renal effects of these steroids are therefore to increase the excretion of K and to reduce the excretion of Na and water. The action of the adrenal cortex on the excretion of salt is thus antagonistic to that of the posterior lobe of the hypophysis. But the action of the adrenal corticoids upon water excretion is not always toward a reduction, it depends upon the degree of salt retention, unless the latter is sufficient to prevent increased excretion of water (reduced reabsorption), this is the effect of the cortical steroids in the normal animal.

The cortical steroids are believed to exert their action, not upon the proximal convoluted tubules (where the great proportion of the minerals of the filtrate are absorbed) but on some more distal part of the nephron. The percentile difference in sodium excretion in an adrenalectomized animal, before and after the administration of cortical hormone, is quite small (about 2 per cent), but since the reabsorption defect is continuous, a profound depletion of the body's sodium stores results with a consequent shrinkage of the extracellular fluid volume. After adrenalectomy the excretion of titratable acid in the urine is diminished, and the alkali reserve of the body fluids reduced (probably as a result of a defect in the mechanism responsible for the excretion of H and K ions for Na ions).

There is evidence that the parathyroid hormone lowers the renal threshold for phosphate. Thyrox-

pituitrin is slight, the diuresis of water ingestion much less pronounced than in later life, and the neurohypophysis of young animals contains only a small fraction of the antidiuretic activity of the adult gland. Before birth the excretory functions, which in after life are served by the kidney, are carried out mainly by the placenta. But the fetal kidney is not entirely functionless for the urea and uric acid concentrations of the amniotic fluid rise steadily to the time of delivery and the bladder of the newborn contains small amounts of hypotonic, acid urine.

¹⁰Though a pronounced diuresis occurs in adrenalectomized animals, strangely enough, the response of the kidney to *ingested* water is defective (ch 59), and water intoxication is readily induced by the administration of water.

ine appears to antagonize the antidiuretic action of the pituitary hormone, water excretion is increased. In hypothyroidism (myxedema) the excretion of water is diminished and the fluid stores augmented.

CRITICISMS AND MODIFICATIONS OF CUSHNY'S THEORY

The question of secretion by the tubular epithelium

From Heidenhain's time to the present innumerable experiments have been performed in attempts to prove or disprove the secretory hypothesis. Only a few of the more important and more recent experimental results can be given.

(a) Marshall and Vickers found that injected phenol red appeared in the plasma in two forms, of which one was *protein-bound* (unfilterable) and the other *free* (filterable). Following injection, more dye appeared in the urine than the quantity of blood which was calculated to have passed through the kidney during the experiment could have contained in filterable form. Part of the dye, it was argued, must therefore have been secreted by the tubules. Also, when the blood pressure was reduced to the point where glomerular filtration ceased and the dye then injected, the kidney contained a higher concentration of dye than the serum or other tissues. Microscopical examination showed deposition of the dye in the cells of the convoluted tubules, it was absent from the lumen. The dye therefore entered the cells from the blood and not simply by re-absorption from the tubules. Sheehan has carried out cognate experiments and arrived at a similar conclusion concerning tubular function.

(b) Chambers and his associates have demonstrated the excretion of phenol red by the mesonephric tubule of the embryonic chick and by the metanephric tubule of a $3\frac{1}{2}$ months human embryo. Employing a micro-manipulative technique the tubules were broken up into sections. Within the first few hours or so of incubation the open ends of the fragmented tubules close and their lumina then become distended with fluid transferred from the surrounding culture medium. Phenol red added to the culture medium was taken up by the tubular cells and concentrated within the cyst-like structures. Cooling the culture to 30°C, CO, H₂S or lack of O₂ stops the process, the tubular epithelium being then permeable in both directions to water and dye.

(c) Evidence for the tubular excretion of creatinine in man has been obtained by comparing its clearance with that of inulin (p 471). Inulin is neither reabsorbed nor excreted by the tubules. Its clearance, which is therefore a measure of glomerular filtration, is considerably lower than that of ingested creatinine. The latter must therefore be excreted by the tubules.

(d) The kidneys of certain fish (e.g. toad fish) possess no glomeruli. Each nephron consists of a segment analogous to the proximal convoluted tubule of mammals. Nevertheless, urea, creatinine, phenol red and such iodine compounds as diodrast and hippuran, when injected appear in relatively large amounts in the urine. Also, as shown by Marshall, the glomerular function of the sculpin can be abolished by phloridzin. The functionally aglomerular kidney, nevertheless, excretes urea, uric acid and injected dyes.

Tubular excretion (of uric acid) is also a prominent feature of renal function in birds and reptiles.

Though the tubules secrete creatinine, H⁺ ions, potassium,¹¹ and a number of foreign substances, e.g., diodrast and dyes, their secretory function, except in so far as it enters into the removal of acid and the conservation of base, appears to be of relatively little importance in man under ordinary conditions of health. This function, however, appears to play a much more important role in chronic renal disease, and is of essential importance in reptiles and birds whose renal tubules secrete large amounts of uric acid.

THE REGULATION OF THE URINARY FLOW

The volume of the urine may be increased by (1) increased rate of filtration, (2) reduced rate of reabsorption, or (3) by a combination of both. Reduction in urinary volume results from corresponding changes in a reverse direction. It should be remembered that a relatively slight percentile alteration in (1) or (2) will cause a relatively great change in the urine volume. If, for instance, the quantity of the reabsorbed fluid remains constant but filtration is increased by only 1 per cent, say, from 100 to 101 liters in twenty-four hours, the urine flow (taking 1500 cc. as the average daily excretion) will be increased by over 60 per cent. A reduction of 1 per cent in the quantity of reabsorbed fluid (filtration remaining constant) will increase the urine volume to a comparable extent. In the human kidney, however, variations in the volume of reabsorbed fluid are of much greater importance in altering the urine flow than are changes in filtration rate, which remains constant under most circumstances.

(1) FACTORS WHICH INFLUENCE THE FILTRATION RATE

¹¹ Berliner has found that potassium is exchanged across the tubular membrane for sodium which is reabsorbed (see p. 459), and that H⁺ and K⁺ ions are excreted by a common mechanism, the utilization of which the two cations are in competition. Suppression of the H⁺ ion exchange by a carbonic anhydrase inhibitor increases the excretion of K⁺ ions.

ATION RATE (a) *Variations in the size of the filtering surface* Richards and Schmidt showed that the number of patent glomeruli in the kidneys of frogs varied from time to time under various nervous and chemical influences (fig. 35.9). Nor is the number of open capillaries in an active glomerulus fixed. At one moment, under a certain condition, only a few loops may be patent, at another moment, under changed conditions, the glomerulus appears as a red ball with all its vessels dilated (Ekehorn). This intermittency of glomerular function, though undoubted in the amphibian kidney, probably does not occur in mammals, the glomeruli in the kidneys of the dog, rabbit and man are continuously active, though it cannot be said that the capillary loops of the glomeruli do not vary to some extent in caliber from time to time.

(b) *Changes in pressure* within the capillary loops which cause a larger fraction of fluid to be filtered from the plasma. Such may result from alterations in caliber of the afferent or efferent vessels (dilatation of the former or constriction of the latter will increase the glomerular pressure), or from changes in the systemic blood pressure. A rise in general arterial pressure, apart from any vascular changes within the kidney itself, will increase both renal blood flow and glomerular pressure. According to Erlanger and Hooker the magnitude of the pulse pressure is also an important factor in renal activity.

(c) *Changes in the renal blood flow* The extent to which a rise in glomerular pressure alone can augment filtration is limited by the fact that as fluid is expressed from the capillary loops the con-

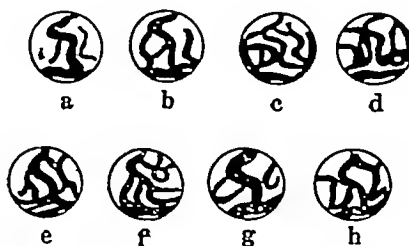


FIG. 35.9 Variations in capillary pathway within a single glomerulus. *a*, 5 minutes before an intravenous injection of 0.1 cc. of 10 per cent glucose; blood flow very slow. *b*, 10 minutes after glucose; blood flow still slow. *c* and *d*, 25 and 30 minutes after glucose; blood flow more rapid and cells less closely packed. *e*, 45 minutes after glucose; blood flow slow, cells densely packed. *f*, 9 minutes after intravenous injection of 0.5 cc. 0.7 per cent NaCl; flow more rapid and cells less dense. *g*, immediately after injection of 0.1 cc. adrenaline 1/100,000; blood flow very slow. *h*, 5 minutes after *g*; blood flow rapid. (After Richards and Schmidt.)

tents of the latter undergo a considerable reduction in bulk. As a consequence, the glomerular blood pressure tends to fall and the osmotic pressure to rise. These effects will, of course, slow the filtration rate. It has been estimated that under ordinary circumstances from about 20 to 25 per cent of the plasma is filtered from the blood in its passage through the glomerulus. The blood within the capillary must therefore undergo renewal at a certain rate in order to maintain filtration at a high level.

(d) *Changes in protein content* and so in the osmotic pressure of the plasma (p. 445)

(e) *Changes in the rate of reabsorption.* In the absence of any other means of facilitating the progress of the filtered fluid through the long and narrow tubules to the renal pelvis, a high pressure within Bowman's capsule would be required (80 mm Hg, according to Brodie). There would then be the need for a glomerular blood pressure sufficient not only for filtration but for the creation of this high intracapsular pressure in order to overcome the resistance in the urinary channels. Reabsorption renders such a pressure unnecessary. In the first place a certain proportion of the filtered fluid need be driven only as far as the proximal convoluted tubule where the reabsorption process commences. The passage of the remainder of the filtrate beyond this to the distal convoluted tubule is facilitated by the withdrawal by reabsorption of the layer of fluid in immediate contact with the wall of the tubule. In the passage of fluid along an ordinary capillary tube the layer of fluid in contact with the wall of the tube is stationary, friction develops between this and the next layer and between successive layers toward the axis of the stream. As a result of the reabsorption process frictional resistance is reduced to a minimum. When, on the other hand, tubular function is impaired the development of a high intracapsular pressure is required in order to overcome the increased resistance and drive fluid along the tubule. Filtration may be materially reduced in consequence.

(2) **THE REGULATION OF REABSORPTION** The chief factors are (a) the rate of flow along the tubules, the faster the flow, the shorter is the time permitted for reabsorption to occur. (b) When sufficient time is allowed, the osmotic pressure of the fluid in the tubules is the factor which brings reabsorption to an end. As a result of the reabsorption from the filtrate of a more dilute fluid than the filtrate itself, concentration of nonthreshold

substances in the tubular fluid and a rise in its osmotic pressure must occur. Reabsorption ceases when the epithelial cells are incapable of performing the work necessary to overcome the osmotic resistance. Back diffusion of low threshold substances tends to increase, however, as the concentration rises. The concentration of the solute in the interstitial fluids, i.e., the fluids on the other side of the tubular membrane, is a factor of equal importance. As this concentration approaches that of the tubular fluid, reabsorption becomes progressively lower and ultimately ceases.

In the reabsorption of a substance from the filtrate there is a maximum rate which may be expressed in milligrams per minute and is conveniently designated by the symbol T_m (maximum transfer). Thus the maximal rate of reabsorption of glucose, or *glucose T_m* , amounts to about 320 mg per minute in man. It is determined by raising the plasma glucose concentration and measuring simultaneously the rate of filtration (inulin clearance method, p. 471) and the urinary excretion of sugar. When reabsorption reaches a maximum, i.e., when the kidney is saturated the excess of unabsorbed glucose of the filtrate appears in large amounts in the urine, subtracting the quantity excreted from the amount filtered gives the quantity of glucose reabsorbed. The figure for the filtration rate multiplied by that for the glucose concentration of the filtrate (i.e., of the plasma) gives the *glucose T_m* .

The reabsorption process is not subject to direct nervous influences, nor, of course, is the purely passive process of filtration. The effects upon renal function exerted by the nervous system are the result of its action upon the circulation and, possibly, upon the activity of certain ductless glands. In the latter connection the results of Theobald and Verney may be cited. These observers found that afferent nerve stimulation caused inhibition of water diuresis of the completely denervated kidney. They suggest that the antidiuretic effect is brought about through pituitary secretion.

THE RENAL BLOOD FLOW

The blood flow through the human kidneys under various conditions has been measured by Smith and his associates by means of the diodrast plasma clearance method. They obtained a value between 1200 and 1300 cc. per minute (both kidneys), or more than one-fifth of the cardiac output. Di-

odrast,¹² an iodine compound, when in low concentration in the plasma, undergoes virtually complete clearance, i.e., it is removed from the plasma in a single passage of the blood through the kidney. This substance is excreted largely by the tubules, only a small fraction (approximately 16 per cent) being filtered through the glomerulus. Knowing the amount excreted in the urine within a unit of time and the concentration in the plasma, it is a simple matter to calculate the quantity of blood which has passed through active renal tissue. By way of illustration let us take some representative figures. Say that 650 mg. of diodrast are excreted in the urine each minute and each cubic centimeter of plasma contains 1 mg., obviously then in the delivery of the 650 mg. of diodrast 650 cc. of plasma, or a little over 1200 cc. of blood, must have passed per minute through the kidneys.¹³ The human kidneys together weigh about 300 grams. A blood flow of around 1200 cc. per minute therefore amounts to about 4 cc. per gram of renal tissue per minute. This figure is in close agreement with that obtained by direct determinations of renal blood flow made upon laboratory animals, the minute flow for the rabbit being given as between 4 and 7 cc. per gram of tissue.¹⁴ If the filtration rate as well as the plasma flow is known, the percentage of the plasma filtered from the blood can be readily calculated from the ratio $\frac{\text{glomerular filtration rate}}{\text{renal plasma flow}}$. Thus, with the usual filtration rate of 125 cc. per minute and a plasma flow of 650 cc. per minute, the plasma in its passage through the glomerular capillaries loses about $(\frac{125}{650} \times 100) =$ about 19 per cent of its bulk. This figure is not constant, but, as indicated above, is altered by several normal and diseased condi-

tions, e.g., the pressure within the glomerular capillaries, the rate of blood flow through them, etc.

Knowing the filtration of plasma filtered, as well as the direction of its change—increase or decrease—in the renal blood flow, changes in the calibers of the efferent and afferent arterioles can be ascertained. Thus, a narrowing of either vessel tends to slow the blood flow through the glomerulus, but constriction of the efferent arteriole increases glomerular pressure and the filtered fraction, whereas a fall in glomerular pressure and a reduced filtration fraction follows afferent arteriolar constriction. Therefore, reduced blood flow, combined with an increase in the filtered fraction, indicates efferent arteriolar constriction, reduction in the filtered fraction and a fall in renal blood flow points to constriction of the afferent arteriole. The first mentioned effect is caused by adrenaline and by renin, the second by exercise. Increased blood flow and a greater filtration fraction indicates afferent arteriolar dilatation. The glomerular circulation is governed largely by the activity of the efferent rather than of the afferent vessel.

Some of the more interesting physiological features of the blood supply to the kidney may be mentioned here. The renal circulation possesses a greater degree of independence in respect to the general circulation than do most other vascular areas. There is little definite information concerning the basis of this autonomy, though the juxtaglomerular apparatus (p. 441) has been suspected of playing a dominant rôle, the controlling mechanism is not impaired by denervation of the kidney. The renal vessels do not share in the reflex vasoconstriction which follows hemorrhage nor in other reflexes initiated from the carotid sinus or aortic arch. Reactive hyperemia (p. 318) which is readily induced in other vascular areas, e.g., skin and cardiac muscle, is not shown by the renal circulation, nor is denervation of the kidney followed by augmentation of the blood flow. Certain fever-producing agents (pyrogens), such as typhoid vaccine, however, cause a pronounced increase in renal blood flow.

¹² Diodone and p-aminohippuric acid are also used for the same purpose.

¹³ In order for a substance to serve as a means of measuring the plasma flow it must not be stored by the kidney nor taken up by the blood cells. All the evidence indicates that in the human subject, at least, these requirements are for all practical purposes fulfilled by diodrast. Diodrast measures only the plasma supplying the parenchyma of the kidney—the *effective plasma flow*—not the small amount supplied to the renal capsule and interstitial tissue.

¹⁴ Van Slyke and associates obtained values in unanesthetized dogs of from 2 to 10 cc. per gram per minute. The blood flow was calculated from the urea N excretion per minute in milligrams (E) and the difference between the urea N contents of renal arterial (A) and renal venous (R) bloods in milligrams per cubic centimeter. Thus,—

$$F = E / (A - R)$$

where F = blood flow in cubic centimeter per minute.

CLEARANCES AND TUBULAR TRANSFER TESTS (SEE CH. 36)

DIURESIS

Substances such as caffeine, sodium sulphate, urea, mercurial salts, acidifying salts, e.g., CaCl_2 , NH_4Cl , and under certain circumstances *digitalis*, act as powerful *diuretics*. Caffeine and the other xanthines, theophylline

and theobromine, exert their diuretic action by depressing the reabsorption of water. They cause no significant effect, as was once presumed, upon the renal blood flow or filtration rate. Caffeine increases the chloride concentration of the urine and lowers the concentration of urea, but the absolute amount of the latter substance excreted is increased (Verney and Winton). The action of *sodium sulphate* is ascribed by Cushny chiefly to reduced reabsorption. Since the epithelial cells are relatively impermeable to sulphate, it increases the osmotic pressure of the fluid within the lumen of the tubules, the work demanded for the reabsorption of the usual quantity of fluid exceeds the capacity of the tubular epithelium. According to Eadie and his associates sulphate also increases the renal blood flow and the filtration rate. *Urea*, and the *ammonium salts* which are reconstituted in the body into urea, act as diuretics in a similar manner. Urea may also increase the filtering surface, more glomeruli becoming active. The *polyuria* of diabetes is due to the high concentration of sugar in the glomerular filtrate, as in the case of sulphate diuresis the osmotic pressure level at which reabsorption ceases is reached sooner than under normal circumstances. The polyuria and glycosuria following the administration of *phloridzin* is due to the paralysis of tubular reabsorptive function. The action of this drug is mainly upon the proximal tubules. *Salycan*, a mercurial diuretic, probably acts also by depressing reabsorption. *Digitalis* has practically no diuretic action in a healthy person nor in a patient suffering from cardiac failure but with no edema. On cardiac edema, but not on other types, the effect is striking, marked diuresis occurs and the edema is reduced. The result is undoubtedly due to the specific action of the drug upon the failing circulation. *Alcohol* is a powerful diuretic, the diuresis resembles water diuresis, it is antagonized by pituitrin.

Water and saline diuresis

The ingestion of large quantities of water by a normal person causes the urine output to increase many-fold. The extra water is eliminated within 4 hours or so, and the greater part of it within 2 hours. The diuresis does not commence until from 30 to 60 minutes after the water has been drunk. Normal or hypertonic saline given by mouth, on the contrary, causes relatively little effect upon the production of urine but when given intravenously causes profuse diuresis. The diuretic effect of the administration of water is much less in anesthetized than in conscious animals. Water diuresis is also inhibited (as a result apparently of the release of the pituitary anti-diuretic substance) by muscular exercise, by afferent nerve stimulation (Theobald and Verney) and in certain emotional states.

After water drinking the urine is very dilute, the specific gravity ranging between 1 000 and 1 002, i.e., considerably lower than that of the plasma. The percentage of chloride as well as the total quantity excreted is greatly reduced. The urea percentage is also reduced, though the absolute quantity excreted may be doubled. The urine in the diuresis, caused by drinking alcohol, shows similar characteristics.

The processes involved in the diuresis following the ingestion of water are complex and not fully understood, the phenomenon has been the subject of numerous investigations. In the experiments of Priestley, from 2 to 3 liters of water were drunk within a few minutes. The diuresis was accompanied by an increase of only about 2 per cent in the water of the blood, and this is quite inadequate to explain the large quantity of urine excreted on the basis of dilution of the plasma proteins. Nor could a rise of 2 per cent in the total blood water account for more than a small fraction of the water ingested. Furthermore, the blood dilution actually passes its maximum and is declining again before the diuresis commences, and if pituitrin is administered at the same time as the water is drunk, diuresis is inhibited, yet the blood dilution is much greater than usual. In these experiments, the results of which have been confirmed in general by others, the blood chloride fell by 5 per cent and the electrical conductivity of the plasma to a corresponding extent, obviously such a reduction in chloride could be only partly accounted for by the blood dilution, and the fall in urinary chloride is evidence that the salt does not escape through the kidney. It appears that the great bulk of the ingested fluid leaves the vessels and is followed, or accompanied by salt which serves to preserve the isotonicity of the tissue fluids. Such a migration of salt would account to a large extent for the fall in blood chloride. The load of fluid is subsequently discharged from the tissues in a slow but steady stream and transported to the kidneys.

From all this it has been clearly established that the diuresis following water drinking is not due to dilution of the plasma colloids by the absorbed water and, as a consequence, to a change in the hydrostatic-osmotic pressure balance (i.e., to an increase in effective glomerular pressure) upon renal function, nor to any direct effect of the water content of the blood upon the kidney. Moreover, if any such factor had this effect, the glomerular

filtration rate should increase. But this does not occur. The delayed onset of diuresis following the ingestion of water is not due simply to the time required for absorption from the intestine, for a similar delay occurs when the diuresis is caused by the intravenous administration of water.

Smirk and associates have also carried out a number of instructive experiments upon water diuresis in man and animals. By means of an ingenious procedure they were able to weigh separately the abdomen and limbs of a human subject or of an intact animal after water drinking. They took the progressive reduction in weight of the abdomen and the increase in weight of the limbs as indices of water absorption from the intestinal tract and water deposition in the tissues, respectively. The weight of the abdomen was found to increase immediately after the water was drunk, and then gradually declined for a period of from 22 to 55 minutes while the leg weight increased. After drinking 1 liter of water the leg weight increased by 1.5 per cent (1 liter was 1.3 per cent of the subject's body weight). This figure agrees

closely with one calculated upon the basis that the absorbed water was distributed generally throughout the tissues of the body. The leg weight gradually fell during the diuresis. The decline was delayed by pituitrin (antidiuretic hormone). Absorption was found to be complete and the dilution of the blood was falling from its maximum before the diuresis commenced (fig. 35 10). The intrarenal factor responsible for the diuresis of water drinking is undoubtedly inhibition of tubular reabsorption, chloride, on the contrary, is reabsorbed more completely. Such urine with its low chloride content resembles that formed by the isolated kidney (that is, one freed from pituitary control). The diuresis following lesions in the region of the pituitary (ch. 57) also resembles the diuresis of water drinking, and like the latter, can be inhibited for several hours by the administration of pituitrin, the chloride content of the urine during the period of inhibition is raised to the normal level. It has been mentioned, page 455, that afferent nerve stimulation inhibits water diuresis. Since the inhibition occurs though the kidney has been de-

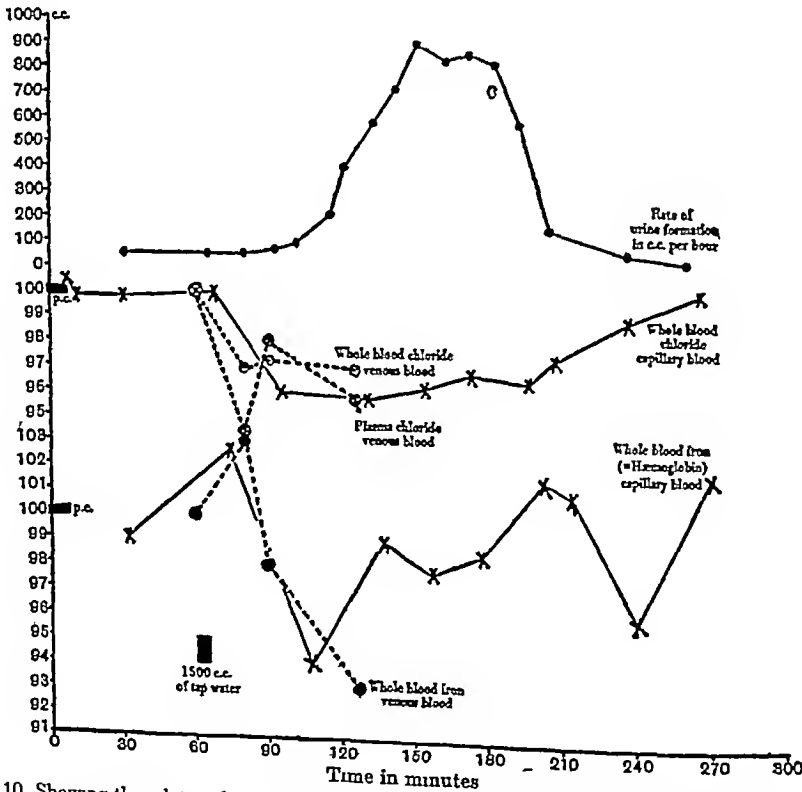


FIG. 35 10 Showing the relationship between the blood changes and the urine production after drinking a large volume of water (After Smirk.)

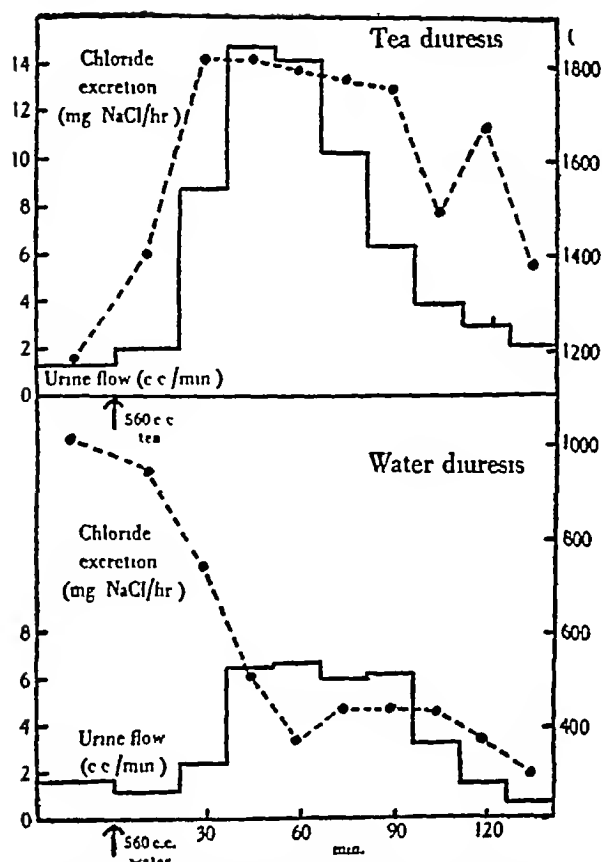


FIG 35 11 The effects of water and of tea drinking (caffeine, p 454) on the urinary excretion of chloride compared (After M G Eggleton)

nervated, it must be brought about through the blood stream

These several observations, especially the characteristics of the urine secreted (extremely low specific gravity and low content of sodium chloride), leave little doubt that the diuresis of water drinking is due to the withdrawal of an antidiuretic influence ordinarily exhibited by the pituitary, what might be called a physiological diabetes insipidus is induced (see also fig 35 11)

Suppression of the antidiuretic action of the pituitary in water diuresis has been clearly demonstrated by Verney and his associates, who have also shown that the liberation of the antidiuretic principle is governed by changes in the osmotic pressure of the blood water, i.e., in the osmotic pressure due to crystalloids, especially sodium chloride, but not to changes in colloid osmotic (oncotic) pressure. The injection of a hypertonic solution of NaCl into the common carotid artery of the dog caused inhibition of water diuresis, resembling that resulting from posterior pituitary extract. The intravenous administration of 10 μ -unit caused a diminution of urine flow equivalent

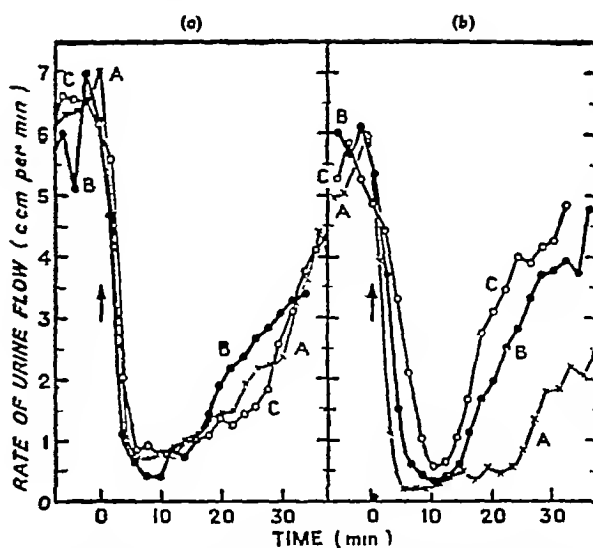


FIG 35 12 (a) Similarity between responses to equivalent increases in the osmotic pressure of the carotid plasma. At the arrow 10.0 cc of 2.50% NaCl was injected into the right carotid in 12 sec (graph A), 10.0 cc. of 15.4% dextrose in 11 sec (graph B), and 2.0 cc. of 8.00% NaCl in 10 sec (graph C), (b) at the arrow postpituitary extract 3.0 μ was injected into the malleolar vein (graph A), 2.0 μ (graph B), and 1.0 μ (graph C) (After Verney)

to that following the intracardiac injection of 21 cc. of 2.50 per cent solution of NaCl (fig 35 12). The effect of the latter was abolished by hypophysectomy, or by ligation of the internal carotid artery. Nor were isotonic solutions effective. Equal amounts of isosmotic solutions of NaCl, glucose or sucrose gave equivalent antidiuretic responses—an indication that the essential factor in the inhibitory effect on diuresis is the rise in the osmotic pressure of the plasma, rather than to any specific chemical effect of sodium chloride. It is concluded, therefore, that structures sensitive to small changes in the osmotic pressure of the plasma water (osmoreceptors) are situated within the area supplied by the internal carotid artery, and that through a change in their activity induced by the osmotically effective substances mentioned, the release of the antidiuretic principle is brought about. A more precise location of these receptors has not been ascertained, but the supra-optic nucleus of the hypothalamus, which is connected with the neurohypophysis by a well-defined nerve tract (hypothalamo-hypophyseal tract, p 784) is the most likely site. It has been estimated that the minimum change in osmotic pressure required to evoke the antidiuretic response is equivalent to about 96 mm Hg, which would correspond to a rise in blood chloride concentration of around 2 per cent. Now, as mentioned

earlier, a fall in blood chloride greater than this follows water drinking, and it is reasonable to assume that such a fall would cause a diuretic effect of the same order as the antidiuresis induced by the rise in osmotic pressure. Liberation of the anti-diuretic principle would thus be suppressed. This latter effect would explain the delay in the onset of the diuresis of water drinking, for a little time must necessarily elapse before all pre-existing hormone had been removed from the circulation.

Release of the antidiuretic hormone may also be brought about through cerebral centers, as well as by afferent nerve impulses from remote parts of the body. The observations of Theobald and Verney have been mentioned, and Connor and Verney have found more recently that water diuresis is inhibited in dogs by emotional states, e.g., apprehension, excitement, anger, etc. The inhibitory effect of exercise upon urine flow appears to be brought about mainly in this way. The irritation or excitement associated with the exercise, rather than the circulatory changes¹⁵ brought about by the exercise itself, appears to be chiefly responsible. It occurs after denervation of the kidney but is not evident after the animal by repetition has become accustomed to the work, unless it is aroused by some loud, ugly sound or is stimulated by a faradic current. The antidiuretic effect of emotion is nearly completely abolished by section of the hypothalamo-hypophyseal tract. It is not mediated through the liberation of adrenaline, for it occurs after the removal of one adrenal and denervation of the other, furthermore, adrenaline is more likely to increase than to diminish the urine flow. Nor is the antidiuretic effect due to a fall in systemic blood pressure for this either rises or remains unchanged. Other effects induced apparently through nerve impulses from psychic centers acting upon the neurohypophysis are the conditioned reflex diuresis which has been demonstrated in ani-

¹⁵ The antidiuresis of exercise in an animal with renal nerves intact may not be entirely hormonal in origin, for the effect is accompanied by a reduction in renal blood flow, blood apparently being diverted to the active muscles. The diminished renal blood flow is associated with constriction of the afferent glomerular arterioles and reduction of the filtration fraction, which indicate that the response to exercise is not due to adrenaline (p. 452). There is a possibility, however, that the blood is diverted through nephrons of the juxtamedullary region which may not extract the material (diuretic or hippurate) used for the determination of the blood flow. In other words, blood shunted through juxtamedullary channels would not be measured.

mals, and the diuresis which can be provoked by suggestion in patients under hypnosis.

REDUCTION IN THE URINARY FLOW

The urinary flow is reduced in muscular exercise (see above), in cardiac failure, as a result of the reduction in renal blood flow (p. 38), in acute inflammatory conditions of the kidney, in surgical shock in some stages of chronic nephritis, in fevers and in dehydrated states, or when the fluid intake is reduced. The effect of pituitrin upon the urine volume has been mentioned.

THE PRODUCTION OF AMMONIA BY THE KIDNEY—HIPPURIC ACID FORMATION

The experiments of Nash and Benedict in 1921 furnished evidence for the renal origin of urinary ammonia. (1) The blood of the renal vein contains 2 or 3 times more ammonia than does the blood of the renal artery. (2) The ammonia of the general circulation is not altered by conditions which increase the ammonia of the urine (p. 461). (3) In animals the blood ammonia does not rise after bilateral nephrectomy, nor in nephritis though the urinary ammonia is often reduced. (4) In diabetic acidosis the blood ammonia is not greater than normal, yet the urinary ammonia is greatly increased.

Urinary ammonia has been thought to be derived mainly from urea and to a less extent from the deaminations of amino acids. However, Van Slyke and his colleagues have shown that glutamine is the chief source. The hydrolysis of glutamine is catalyzed by an enzyme in the kidney—*glutaminase*. Amino acids furnish a smaller proportion, but none is derived from urea. They found that all the urea removed from the blood by the kidney was excreted unchanged, whereas the glutamine so removed accounted for at least 60 per cent of the urinary ammonia.

The kidney (as well as the liver) conjugates glycolic and benzoic acid with the production of hippuric acid (benzoyl-glycine). This is a detoxicating process which probably occurs in the distal tubules.

THE REACTION OF THE URINE AND THE RÔLE PLAYED BY THE KIDNEY IN THE REGULATION OF THE ACID-BASE BALANCE

Regulation of the acid-base balance of the body fluids and the maintenance of the normal concen-

trations of electrolytes are two of the most important functions of the kidney

The acidity of the urine may be expressed in two ways, either with regard to (a) *its concentration in hydrogen ions*, i.e., its true acidity, or (b) the *total quantity of free acid* present as determined by titration with N/10 sodium hydroxide using neutral red or phenolphthalein as indicator. This measurement, which is expressed as the number of cubic centimeters of N/10 alkali used for the titration, is called the *titratable acidity*.

(a) THE pH OF NORMAL URINE usually lies between 5.0 and 7.0 with a mean of 6.0. The extremes in health are 4.7 and 8.0. The pH of urine is determined mainly by the proportions of the di-basic (alkaline) and mono-basic (acid) phosphates which it contains. The kidney receives blood at a pH of about 7.40 and forms urine with a pH of around 6.0. It accomplishes this partly by altering the proportions of the acid and alkaline phosphates. The ratio of these in the plasma is $\frac{\text{NaH}_2\text{PO}_4}{\text{Na}_2\text{HPO}_4} = \frac{1}{5}$ whereas in urine with a pH of 6 the

ratio is more than reversed, being $\frac{\text{NaH}_2\text{PO}_4}{\text{Na}_2\text{HPO}_4} = \frac{9}{1}$

When the urine is at its greatest acidity the ratio is 50:1. By this device a large amount of the body's fixed base (Na chiefly, but also K, Mg and Ca) is conserved. The change from di-basic to the mono-basic salt probably takes place during the passage of the urine through the distal tubule (see fig. 35.13). Wearn and Richards found that the glomerular filtrate of frogs was alkaline in reaction whereas the

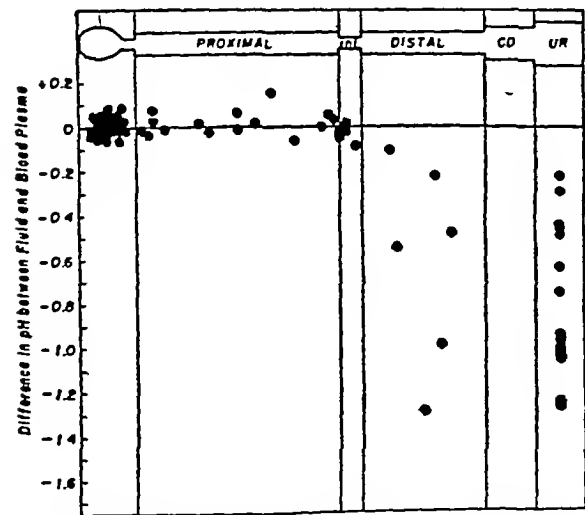
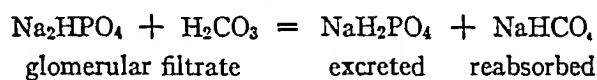


FIG. 35.13. Diagram to illustrate the site of acid production in the kidney. (After Pitts)

urine was acid. Cushny found that in dogs the injection of di-sodium (alkaline) phosphate increased the acidity of the urine. A portion of the sodium of the injected salt was retained while the resulting acid phosphate was excreted. It was also observed that when the di-basic salt was injected into an animal with one ureter partially obstructed (to increase absorption from the tubules) the urine from that side was more acid than was the urine coming from the free ureter. The injection of such substances as sodium chloride or sodium sulphate was found to reduce the acidity of the urine since they caused diuresis, the rapid flow of urine through the tubules curtailing the conversion of di-basic phosphate to the mono-basic salt.

One or other of two mechanisms has been suggested for the increase in the proportion of acid phosphate in the urine. Either conversion takes place in the tubules as shown below, the bicarbonate being reabsorbed. Or, the alkaline salt may be preferentially reabsorbed leaving the acid phosphate to be excreted (see fig. 35.14).



Pitts and his associates offer evidence for an additional and apparently more important mechanism in the elimination of acid by the kidneys. They found that much more acid per unit of time was excreted in the urine than could be derived from the glomerular filtrate. The excess acid must be secreted by the tubules. But it is not definitely established whether its elimination is as molecular acid, e.g., HCl, or by ionic exchange. In the former instance the HCl would combine with base in the

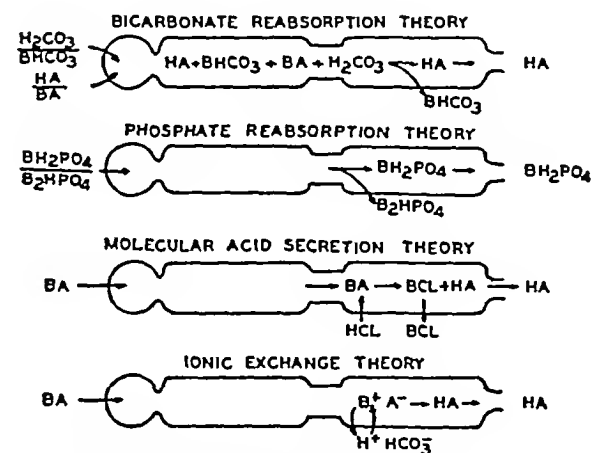


FIG. 35.14. Illustrating theories of acid production by the kidney. (After Pitts)

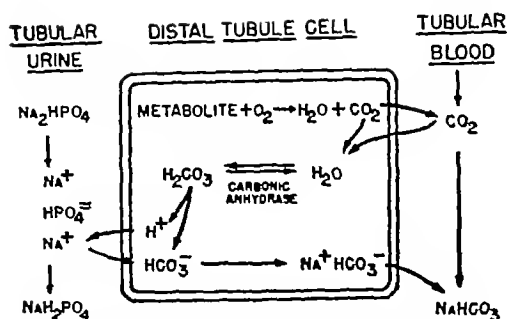
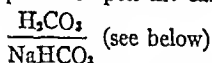


FIG. 35 15 Showing the cellular mechanism, according to the ion-exchange theory, for acidification of the urine (After Pitts)

tubules to form NaCl , which would then be reabsorbed. In an ionic exchange mechanism, which is the most probable, H^+ ions in the plasma derived from H_2CO_3 would be exchanged for ions of fixed base, chiefly Na^+ , across the tubular membrane. The high concentration of carbonic anhydrase (pp 132 and 398) and the depressing effect upon the renal excretion of acid by sulfanilamide, which inhibits the action of carbonic anhydrase, lends support to the existence of an ionic exchange mechanism. In either case, however, the end results would be identical (see figs 35 14 and 35 15). In alkalosis the urinary pH rises and may become alkaline, then relatively large amounts of bicarbonate are excreted. The pH of the urine is thus mainly dependent upon the carbonic acid-bicarbonate ratio



The pH and volume of the urine in health and under usual conditions vary together and in the same direction, a urine of small volume being usually of a lower pH than one of larger volume. A urine of average acidity (pH 6.0) may have its pH raised to 7.0 by maximum diuresis but rarely higher than this.

(b) **TITRATABLE OR TOTAL ACIDITY** By this is meant the quantity of acid in excess of basic substances (i.e., the free acid) present in the urine. The greater part of the titratable acid is phosphoric in the form of the mono-basic (chiefly sodium) salt. A small part of the titratable acid is due, however, to free organic acids, uric, lactic, hippuric, etc. The usual amount of titratable acid in a 24-hour specimen of urine lies between 200 and 400 cc. of $\text{N}/10$ standard acid. Its amount, however, depends in health upon several conditions and may be as low as 100 cc. or as great as 600 cc. In severe diabetes it may be 2 or 3 times the latter figure, as a result of the excretion of large quantities of acetoacetic and β -hydroxybutyric acids (acetone bodies).

The total acidity is determined by titrating the urine with $\text{N}/10$ sodium or potassium hydroxide, the result being expressed as the number of cc. of $\text{N}/10$ alkali required to bring 100 cc. of urine to near the neutral point, the precise reaction attained depending upon the indicator used. When the titration is carried out to the pH of blood (7.4) the result obtained has a greater physiological significance than if the urine were exactly neutralized (pH 7.0). In Henderson and Palmer's method neutral red, which turns a brownish red at the normal blood reaction, is used as indicator. When the urine is titrated to the reaction of the blood, the procedure obviously reverses the change in reaction which the kidney had effected during the formation of the urine from plasma. The quantity of alkali added is therefore a measure of the quantity of fixed base which the kidney has saved to the body by the conversion (chiefly) of the di-sodium phosphate to the corresponding acid salt. Frequently, phenolphthalein is used as the indicator in titrations of urinary acidity (Folin's method) but since this indicator turns color (colorless to pink) at a pH of 8.2, it is clear that the result will be considerably higher than when neutral red is employed, and urine which is actually neutral or alkaline, or even blood plasma itself will then have a titratable "acidity."

BASE CONSERVATION BY THE KIDNEY

The stores of base in the body fluids can be replenished only from without. The kidney plays an important rôle in preventing the depletion of these stores. In the conversion of the di-basic to the mono-basic phosphate, mentioned in the preceding section, one molecule of NaHCO_3 is returned to the body's store of alkali for each molecule of the alkaline phosphate converted in the tubules to the acid salt. Or, put in another way, about 40 per cent of the base bound to phosphoric acid in the plasma is saved by the conversion of the di-basic phosphate to the mono-basic form.

Another means of conserving base is the retention of part of the base combined in the blood with organic acids and the excretion of the latter in the free state, this though of little importance in health is of great importance in disease. Organic acids may pass into the urine in three ways: (a) in combination with fixed base (urates, hippurates, lactates, etc.), (b) in combination with ammonia, or (c) free. In highly acid urines according to Henderson and Palmer uric acid is 91 per cent free, lactic 12 per cent and hippuric 8 per cent. The organic acids acetoacetic and β -hydroxybutyric formed in excess in diabetes though combined with fixed base to the extent of 100 per cent in the plasma with consequent reduction in the

alkali reserve are excreted free in considerable quantities (acetoacetic, 5 to 11 per cent, β -hydroxybutyric, 20 to 55 per cent) Any acid excreted in the free state obviously represents a clear saving of the body's alkali Palmer and Van Slyke state that the measurement of the increase in the excretion of total organic acids in diabetes is a reliable approximated guide to the quantity of acetone bodies The total quantity of organic acid (bound and free) excreted daily by a healthy man is between 240 and 600 cc of N/10 organic acid (average of 6.0 cc per kg. of body weight)

Still another method by which fixed base is saved by the kidney is in the manufacture of ammonia¹⁶ (p. 458) for the neutralization of acids The kidney separates certain acid radicals from the fixed bases with which they were combined in the plasma, joins them to ammonia, and excretes them as ammonium salts In order therefore to arrive at a true estimate of acid excretion and the base saved to the body, the quantity of the ammonia excreted should be added to that of the titratable acid (titration to 7.4) The acid + ammonia value represents what may be termed the *base economy* exercised by the kidney In health the daily excretion of ammonia amounts to from 300 to 500 cc of N/10 ammonia In severe diabetes it may be ten times this value The hydrogen ion concentration, titratable acidity and ammonia of the urine usually vary together and in the same direction, though there is no exact proportional relationship All three are determined by the relative concentrations of basic and acid radicals in the plasma The value of the acid + ammonia excretion may therefore be employed as an approximate index of the lowering of the alkali reserve in such conditions as diabetes The *ratio* of ammonia excretion to titratable acid in health is usually 1 or over ($\text{NH}_3/\text{titratable acid} = 1 \text{ to } 2.5$) In moderate degrees of kidney disease in which the urea concentrating function is impaired the ratio is low (0.1 to 0.7) owing to the fact that the power of the kidney to form ammonia is depressed while the excretion of acid is little altered When however the renal insufficiency becomes extreme, the ratio may be higher than the impairment of renal function would lead one to expect The rise in the ratio is due, not to an increase in the output of ammonia, but to a reduction in the excretion of acid In the nephrotic

type of nephritis in which there is no impairment of urea excretion the production of ammonia is not interfered with and the ratio is normal

Fixed base enters the urine other than in combination with phosphoric acid Sulphuric and hydrochloric acids formed during protein catabolism differ from phosphoric acid in that they are not excreted in the free state except in infinitesimal amounts but carry with them their full base equivalent, appearing in the urine as the neutral salts of either fixed base (NaCl , Na_2SO_4) or of ammonia (NH_4Cl , $(\text{NH}_4)_2\text{SO}_4$)

The free carbonic acid in the urine does not vary in amount with changes in urinary pH The bicarbonate content of the urine, on the contrary, varies directly with the pH, in alkaline urines its quantity is large (up to 8 grams per liter), whereas in highly acid urines its quantity is negligible (0.02 gram per liter) When, therefore, the alkali reserve of the blood is normal or low, the urine is acid and practically no base is lost as bicarbonate When, on the other hand, the bicarbonate of the plasma is above normal the urine is alkaline and large quantities of base pass into the urine as bicarbonate Since the free carbonic acid remains unchanged at varying urinary pHs while the bicarbonate content varies, the ratio $\text{H}_2\text{CO}_3/\text{NaHCO}_3$ in the urine varies with changes in pH The ratio in urine on the acid side is from $\frac{1}{10}$ to $\frac{1}{5}$, whereas in neutral or alkaline urines it is from $\frac{1}{5}$ to $\frac{1}{10}$ The small quantities of free carbonic acid and bicarbonate present in acid urines render these "buffers" of negligible importance in determining the urinary pH as compared with the alkaline and acid phosphates In alkaline urine the reverse is true, the carbonic acid-bicarbonate ratio being of primary importance

The influence of starvation, diet, and of acid and alkali ingestion, upon urinary acid

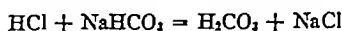
Starvation increases the titratable acidity and ammonia, and lowers the pH of the urine After exhaustion of the glycogen stores increased protein breakdown occurs with the production of phosphoric and sulphuric acids As a result of the incomplete combustion of fat acetoacetic and β -hydroxybutyric acids are formed and excreted

Diet Cereals, meat and fish increase the titratable acid and ammonia of the urine and lower the pH Most fruits (apples, pears, peaches, lemons, oranges, grapes, figs and raisins) contain the salts of organic acids, the acid radical undergoes oxidation in the body, the alkali being liberated These fruits therefore tend to raise the alkali reserve of the plasma and to reduce the acidity of the urine Plums, prunes and cranberries, on the contrary, increase urinary acidity since they contain benzoic and quinic acids which are converted to hippuric acid in the kidney The herbivora normally pass alkaline urine, though if starved and so forced to

¹⁶ Pitts postulates that undissociated ammonia diffuses into the urine and then forms ammonium ions by combination with H-ions

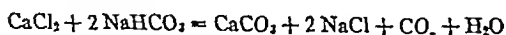
use their body protein or if fed upon flesh, the urine becomes acid. The pH of the urine of men upon a vegetarian diet is from 5.30 to 7.48 with a mean of 6.64 instead of the usual mean of 6.0.

Ingestion of acid and alkaline chemical substances
When acids such as sulphuric or hydrochloric, or acidifying salts (CaCl_2 , NH_4Cl , $(\text{NH}_4)_2\text{SO}_4$) are taken by mouth in large amounts, a reduction in the alkali reserve of the plasma, an increase in the total excretion of phosphate, chiefly of the mono-basic salt, a rise in titratable acidity and a fall in pH of the urine result. The urinary ammonia increases. The following reaction results from the absorption of hydrochloric acid



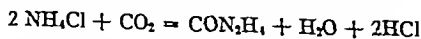
A proportion of the Cl^- radical is excreted in the urine as NaCl and so entails an equivalent loss of fixed base. Some excretion of acid over basic substances is effected by the excretion of a greater proportion of the mono-basic phosphate. This factor, however, is limited since at the usual reaction of the urine 90 per cent of the di-basic phosphate has already been converted to the mono-basic form. Also, the body's available stores of phosphate after a time become exhausted, the excretion of acid by this means must then come to an end. It has been pointed out that normally a most important factor in the elimination of acid radicals and the conservation of fixed base is the production of ammonia by the kidney, after the ingestion of HCl a part of the Cl^- is excreted as the ammonium salt.

Similar changes to those just described follow the ingestion of CaCl_2 or of NH_4Cl in large doses. In the case of calcium chloride ingestion Cl^- displaces HCO_3^- from the bicarbonate of the body fluids according to the following equation



The NaCl and CaCO_3 are excreted. There is thus a net loss of base from the body, a severe acidosis results. The calcium is excreted mainly in the feces, though it is possible that a proportion is also deposited in the bones, since calcium in the form of the chloride when fed to dogs cannot all be recovered from the excreta. As in the case of HCl ingestion, Cl^- is also excreted in the urine combined with ammonia.

The acidifying effect of NH_4Cl is probably due to the conversion of NH_4 to urea, the Cl^- radical being freed. When 20 grams or so of NH_4Cl are administered the urinary excretion of urea and of sodium chloride rises abruptly. The reaction is shown as follows



The HCl then reacts with bicarbonate of the plasma with the production of sodium chloride and carbonic acid. The passage of the large quantities of base into the urine when calcium chloride or ammonium chloride

are administered requires the excretion of large quantities of water. These two salts therefore act as powerful diuretics and agents for the removal of edema fluid. When large quantities of acid sodium phosphate are ingested, though there is a great increase in the titratable acidity of the urine, no change in ammonia excretion occurs. Since this salt can be excreted as such there is no call apparently for the production of ammonia for the conservation of base such as occurs after the ingestion of hydrochloric and other acids.

The effects of the *administration of alkalis* are, as might be expected, the reverse of those following the ingestion of acid substances. Large doses of bicarbonate augment the alkali reserve, reduce the excretion of total phosphates, and increase the proportion of those in the di-basic form. The excretion of the titratable acid plus ammonia falls while the bicarbonate content and pH of the urine rise. Though large doses (45 grams daily) of bicarbonate are required to render the urine alkaline and keep it so, a wave of urinary alkalinity will result from the administration of much smaller doses. Palmer and Henderson make use of this fact in their *bicarbonate tolerance test* for reduction in alkali reserve. Four grams of sodium bicarbonate when administered to a normal person cause a perceptible rise in urinary pH, whereas larger doses are required to produce a similar effect in one with a lowered alkaline reserve. A similar clinical test has been devised by Sellards. These tests are of value in some conditions, but in nephritis the power of the kidney to excrete bicarbonate may be impaired, under which circumstances the test would indicate a lower alkaline reserve than actually existed. In diabetes also the test may indicate acidosis in its absence. Salts of organic acids—acetates, citrates, lactates—are also alkalinizing in their action, the organic acid radical is oxidized, the base joining with H_2CO_3 to form bicarbonate.

The "alkaline tide" of the urine

Leathes found that the urine passed a short time after rising in the morning was less acid than that formed during sleep. This change in urinary reaction is spoken of as the *morning alkaline tide*. (The urine does not necessarily become alkaline but may only become less acid—an elevation in pH or decreased titratable acidity.) The greater acidity of the night urine over the urine of the early waking hours was ascribed by Leathes to depression of the respiratory center during sleep and the retention of carbon dioxide. The increased pulmonary ventilation which occurred after rising supposedly removed carbon dioxide, the effect upon the acid base balance being reflected in the reduced acid excretion by the kidney.¹ Leathes found that samples of alveolar air taken immediately upon awakening had a higher percentage

of carbon dioxide (6.8 per cent) than those taken an hour or so after rising (CO_2 , 5.7 per cent). It has been shown by Hubbard, however, that merely awakening causes a rise in urine pH which may be actually higher than after getting out of bed. Now, it is well known that a change in the reaction of the urine towards alkalinity may be demonstrated by voluntary forced breathing. (If the forced breathing is continued for some time the urine may become alkaline to litmus, the bicarbonate content is increased, the acid + ammonia is reduced and acetone bodies may appear.) Nevertheless, since the morning alkaline tide is more pronounced immediately after waking than later, when the exercise of dressing, etc., should increase the pulmonary ventilation, the mechanism underlying the fall in urine acidity is not so simply explained. It is associated in some unexplained way with respiratory changes incident to the return of consciousness.

An alkaline tide also occurs within an hour after a meal—*postprandial alkaline tide*. The titratable acidity and ammonia excretion are reduced and the urinary pH rises.

A rise in the CO_2 tension of the alveolar air was first shown by Higgins to occur almost immediately after taking food. Dodds found the rise to amount to from 2 to 6 mm Hg, and Van Slyke, Stillman and Cullen observed a corresponding rise in the plasma bicarbonate. These changes—reduced excretion of acid by the kidney and retention of carbon dioxide—have been generally attributed to compensatory adjustments in the acid-base balance resulting from the loss of hydrochloric acid in the gastric juice. Dodds also reported a fall in alveolar CO_2 tension in from one to two hours after a meal, which was put down to the secretion of alkali in the pancreatic juice. The view that gastric secretion is the cause of the alkaline tide of the urine was supported apparently by the fact that it was most pronounced after the ingestion of meat—a powerful gastric stimulant—whereas it was much less likely to appear after a meal of carbohydrate or fat. Nevertheless, there are many difficulties in the way of accepting this view. A well-marked alkaline tide following meals has been observed in subjects of achlorhydria.

¹⁷ Brunton points out that the acid output per minute, that is, the titratable acidity value multiplied by the urine volume (cubic centimeter) per minute is a more reliable index of acid excretion than simply the titratable acidity per unit volume of urine. The latter may be reduced as a result of diuresis without the actual quantity of acid secreted in a given time being altered.

Brunton and Wilson, on the other hand, have produced good gastric secretion in subjects by means of a meal of Bovril without any decrease in the total excretion of acid by the kidney. Brunton and Israels also report that a rise in alveolar CO_2 tension of an increase in the quantity of CO_2 expired per minute does not, as a rule, occur when a profuse secretion of gastric juice is induced by the injection of histamine. A fall in CO_2 tension after a meal coincident with the calculated time of pancreatic secretion was not observed. Such observations are strongly opposed to the idea that the rise in alveolar CO_2 tension following meals is a consequence of gastric secretion. It is more likely connected in some way with the drowsiness and the reduction in pulmonary ventilation which occurs at this time. Main attributes it to the removal of the stimulating effect of the hunger contractions on the respiratory center. It is also improbable, on theoretical grounds, that the loss of HCl is a factor, since pancreatic secretion follows so closely upon gastric secretion that one effect would be expected to balance the other. These remarks also apply to the postprandial alkaline tide of the urine. Possible additional factors in the production of the latter are excess of basic over acid radicals absorbed from the food itself or the sparing effect of the meal upon protein catabolism. In some instances the decreased urinary acidity is simply due to diuresis, the acid content per unit volume of urine being reduced without any change occurring in the total acid excretion.

THE VOLUME OF THE URINE

The quantity of urine excreted in 24 hours by a healthy adult under ordinary circumstances is between 1000 and 1800 cc. The normal kidney may, depending upon the requirements at the time, eliminate per hour as little as 25 cc of fluid or as much as 1200 cc. Usually from 40 to 60 per cent of the total fluid intake of the 24 hours is excreted by the kidneys, but the quantity of urine formed must necessarily vary inversely with the quantity of fluid eliminated by other channels,—lungs, skin and bowels. Diet, the quantity of fluid drunk, environmental temperature and humidity, posture, exercise, mental excitement, weight, age and sex, are among some of the physiological factors which influence the urine volume. The output is greater upon a high than upon a low protein diet, the end products of protein metabolism exerting a diuretic effect. High temperature, espe-

cally if accompanied by visible sweating, reduces the urine volume. Muscular exercise has a similar effect. The volume is greater when the subject is lying down and awake than when upright, as a result of reflex vasoconstriction (of afferent glomerular arterioles) and also probably if the erect posture is maintained for long, to the leakage of fluid from the circulation in dependent parts and, in consequence, increase in concentration of the plasma proteins. The urine volume is reduced below the normal in cardiac failure, in fevers, in acute nephritis and in stages of chronic nephritis. The volume is markedly increased in diabetes insipidus (p 802) and in diabetes mellitus (For the action of diuretics see p 454.)

Young children, especially infants, excrete for their weight 3 or 4 times more urine than do adults. The ability of the kidney to concentrate the urine does not develop fully until several months after birth, the urine therefore is isotonic or hypotonic, as well as being of large volume.

The urine of the day (8.30 a.m. to 8.30 p.m.) is normally from 2 to 4 times greater in amount than that secreted during the night (8.30 p.m. to 8.30 a.m.) This ratio holds even though the quantity of fluid drunk during both periods is the same. In normal night workers the ratio is reversed. An increase in the night urine of a day worker to an amount approaching that of the day period should suggest renal insufficiency (*nycturia*).¹⁸ In many cases of renal disease the night urine exceeds the day urine in volume. The terms *polyuria* and *oliguria* are applied respectively to the excretion of supernormal and subnormal quantities of urine. Total cessation (suppression) of urinary excretion is termed *anuria*.

The influence of the pituitary upon the urine volume is dealt with in chapter 57 (see also pp 466 and 477).

SPECIFIC GRAVITY OF THE URINE

The specific gravity of the urine of adults under the ordinary conditions of health varies between 1.015 and 1.025 and inversely with the urine volume. A urine of large volume usually has a lower percentage of total solids and consequently a lower specific gravity than one of small volume. Upon a very high fluid intake a urine with a specific gravity of 1.001 may be passed. The urine excreted upon a low fluid intake, or when much fluid is lost

through other channels, skin and bowels, may have a specific gravity as high as 1.030. In *pathological conditions* this relationship between specific gravity and volume does not necessarily hold, for large quantities of urine with a very high specific gravity (1.040 or more) are eliminated in diabetes mellitus with much sugar in the urine, whereas in chronic nephritis a urine of small volume and of low specific gravity (1.010) may be excreted. Though the fluid intake be rigidly restricted a kidney whose function is seriously impaired may be unable to concentrate the urine sufficiently to raise the specific gravity above 1.010 or so. Furthermore, when the fluid intake is high the production of urine of very low specific gravity (1.008 or lower), a feat readily accomplished by the normal kidney, is beyond the powers of a seriously diseased kidney. Variability of function according to the needs of the moment is one of the most pronounced characteristics of the normal kidney. The diseased kidney on the contrary is unable to vary the specific gravity of the urine over a wide range, its ability to alter the volume of the urine is also very greatly impaired. The specific gravity of the urine in health is directly proportional to the concentration of its dissolved solids (chiefly chlorides and urea). The specific gravity of the infant's urine, as mentioned above, is low owing to the immature state of the infantile kidneys—low concentrating ability.

A rough method of determining the total urinary solids is to multiply the last two figures of the specific gravity of 2.66 (Long's coefficient). This gives the quantity of total solids in grams per liter. This method is not reliable for some pathological urines, since different substances vary in the extent to which they contribute to the specific gravity of a solution, it is clear that should the urine contain abnormal constituents (albumin, sugar) or should the proportions of its normal constituents be altered greatly, the relationship between total solids and specific gravity just expressed will not hold. Less than 1.5 grams of NaCl, for instance, will produce as great a rise in specific gravity of the urine as nearly 4 grams of albumin. The latter unless present in relatively large quantities will therefore have little effect upon the urinary specific gravity. The following table (after Albarran) gives the amounts of several substances which are required to be added to a liter of urine in order to raise its specific gravity 0.001.

	grams
Urea	3.6
Glucose	2.7
Acid sodium phosphate	3.8
Sodium chloride	1.47
Albumin	3.9

¹⁸ An increase in the volume of night urine above 500 cc. and with a specific gravity less than 1.018 is usually defined as *nycturia*.

COMPOSITION OF THE URINE

In the following table are given figures for the average amounts of inorganic and organic solids in 100 cc. of urine of a healthy adult on an ordinary mixed diet. The total solids amount to from 40 to 50 grams per liter, the urea constituting about one half of this.

I. Inorganic constituents

	grams per 100 cc.
(a) Chloride expressed as NaCl	9.0
(b) Trisphosphate expressed as P_2O_5	2.0
(c) Total sulphur expressed as SO_4	1.5
(d) Sodium expressed as Na ₂ O	4.0
(e) Potassium expressed as K ₂ O	2.0
(f) Calcium expressed as CaO	0.2
(g) Magnesium expressed as MgO	0.2
(h) Iron	0.003

II. Organic constituents

A. Nitrogenous

Urea	25.0 grams containing approximately, 10.0 grams nitrogen
Ammonia	0.6 gram containing approximately, 0.4 gram nitrogen
Lactic acid	0.6 gram containing, approximately, 0.2 gram nitrogen
Creatinine	1.5 grams containing, approximately, 0.5 gram nitrogen
Undetermined nitrogen ¹	0.6 gram
Total nitrogen	11.7 grams

The urea, ammonia, and the inorganic and ethereal sulphates (pp. 637 and 638) vary with the protein content of the diet whereas the creatinine and neutral sulphur, particularly the former, are little affected by the level of the protein intake (see also table 55, p. 637).

B. OTHER ORGANIC SUBSTANCES PRESENT IN SMALL QUANTITIES OR MERELY IN TRACES

- Pigments, e.g., urochrome and urobilin, (traces of the latter only)
- Sugar, minute amounts (2 to 3 mg. per cent) of fermentable sugar are found in the urine in 50 per cent of normal persons after the noon meal (Harding and associates)

- Fatty acids, aceto-acetic and β hydroxybutyric acids
- Carbonates, bicarbonates and free carbonic acid
- Mucin and mucin like substances
- Diastase

Phosphorus is found in the urine largely combined with sodium as the *mono sodium* and *disodium phosphates* NaH_2PO_4 and Na_2HPO_4 , and to a much less extent with potassium to form the corresponding salts of this mineral. About three-quarters of the total urinary phosphorus exists in these forms. The remaining quarter or so of the total phosphate is present as salts of the alkaline earths, calcium and magnesium. These earthy phosphates are insoluble in alkaline urine and for this reason are precipitated (phosphate deposit) when, as a result of bacterial action the urea of urine which has stood for some time after voiding becomes decomposed with the formation of ammonia. Infection of the genito-urinary tract may for the same reason produce an alkaline urine which when voided is clouded by precipitated phosphates. In perfectly healthy persons urine which, as a result of a diet rich in alkaline carbonates, is alkaline in reaction may be cloudy when passed owing to the production of the insoluble phosphates of calcium and magnesium $Ca_3(PO_4)_2$ and $Mg_3(PO_4)_2$. The condition is called *phosphaturia* and is quite physiological.

The total phosphate content of the urine is *increased* under several conditions,—lowered alkali reserve, starvation, high protein diet, severe muscular exercise, in gout and following the administration of large doses of parathyroid extract. It is increased by such diuretics as sodium nitrate and sulphate, and by calcium and ammonium chlorides, but not in water diuresis. The urinary P is low in the morning but rises in the later hours of the day, reaching a maximum in the night urine. Though the great part of the urinary P is derived from the inorganic P of the plasma, a small part (6 per cent) is organic (glycerophosphates, hexose phosphates, nucleotides). The kidney is rich in *phosphatase* (p. 867) yet it does not seem that any considerable part of the inorganic P of the urine results from the action of this enzyme upon organic phosphorus compounds of the plasma. The urinary P is *reduced* in pregnancy, when the alkaline reserve is high, in renal insufficiency, tetany and in certain bone diseases. Insulin causes a marked fall in the urinary P for the first six hours following the injection, and a rise for the next six hours. The work of Brull and Eichholtz suggests that the pituitary or hypothalamic region of the brain is concerned in the excretion of inorganic P by the kidney. This was suppressed by removal of the gland.

¹ This includes the nitrogenous substances of the urine other than those enumerated above—amino acids, hippuric acid (as sodium hippurate), allantoin, uric acid, etc.

or by injury to the tuber cinereum, yet no corresponding change in the level of plasma P (organic or inorganic) occurred ²⁰

Sulphur occurs in the urine in three forms (a) *inorganic sulphates*, (b) *etherical sulphates*, and (c) *neutral sulphur* (see p 637 and 638)

Urochrome (Thudichum) is the normal urinary pigment. Its chemical constitution and the nature of its precursor are unknown. Its output was found by Drabkin to be remarkably constant from day to day under the ordinary conditions of health. The output is independent of diet but bears a relation to the level of the basal metabolism. It is therefore endogenous. In fevers, hyperthyroidism, and as a result of elevating the metabolic rate by the administration of thyroxine or adrenaline its output is increased. The output of urochrome is also increased by tissue breakdown, by starvation or by the administration of acids. Thyroid extirpation reduces its elimination below the normal level in

²⁰ Except in the alkalosis and hyperphosphatemia of pyloric stenosis when the urinary P is increased

animals, and to the normal level in subjects of exophthalmic goiter

The *chloride* of the urine (chiefly in the form of NaCl) is derived almost solely from the chlorides of the food. The concentration of chlorides in normal urine varies from 0.5 to 2.0 grams (as NaCl) per 100 cc. as compared with about 0.6 per cent in the plasma. Chloride is a high threshold substance and in prolonged starvation or upon a salt-free diet, though the urine may be practically chloride free, the chloride of the blood remains at or but slightly below the normal level. When chloride is lost through vomiting, the urinary chloride is reduced, the Cl lost in the vomitus being derived from the blood NaCl, results in an excess of base (Na) being left in the plasma, this combines with H_2CO_3 to increase the bicarbonate service.

The experiments of Verney and Starling demonstrated quite clearly the influence of the pituitary upon the excretion of chloride (p 456)

CHAPTER 36

THE PATHOLOGICAL PHYSIOLOGY OF KIDNEY DISEASE

MICTURITION

Classification of kidney diseases

A logical classification of kidney disease would be one based upon the nature (inflammatory or degenerative) and the site (glomeruli, tubules or vessels) of the renal injury, and which would permit a correlation to be established, through our knowledge of renal physiology, between such injury and the signs and symptoms of the disease. Such an ideal is impossible of attainment, for though a certain part of the renal structure at first bears the brunt of the disease, there is a tendency for other parts of the kidney to become progressively involved, if not in anatomical changes, at least in function. For this reason, as well as from the fact that extrarenal factors (e.g., cardio-vascular) may become disturbed to react upon renal function, the signs and symptoms of chronic kidney disease are extraordinarily diverse, mixed forms of nephritis rather than pure types are common. Nevertheless, a classification based upon the nature of the *primary* pathological process (inflammatory, degenerative, or arteriosclerotic) and the type of renal structure *predominantly* affected will be given. Where possible an attempt will be made to explain signs and symptoms by the application of physiological principles. There are three main types of renal disease.

(1) *Glomerulo-nephritis* is essentially inflammatory in nature, and the glomeruli are primarily affected.

(2) The *nephroses* are the result of a degenerative process chiefly involving the tubules.

(3) *Arteriosclerotic kidney disease*. The large or the small renal vessels are the seat of sclerosis.

GLOMERULO-NEPHRITIS

Glomerulo-nephritis is seen in an acute and a chronic form.¹ Though the chronic type is believed to be nearly always initiated by the acute disease, the latter may have been so mild as to have escaped notice. So, chronic glomerulo-nephritis often appears to have arisen *de novo*. The

¹ Clinically a subacute form—a transition stage between the acute and chronic forms—is recognized.

acute disease, and so the chronic form into which it merges, is thought to be usually due to the toxin of some streptococcal or pneumococcal infection, e.g., sore throat, scarlet fever, influenza, etc.

ACUTE GLOMERULO-NEPHRITIS

In the acute disease the glomeruli show inflammatory changes which vary in intensity in different cases. When the disease is severe, the following histological picture is seen. Great numbers of the glomerular capillaries are dilated and filled with a coagulated exudate containing a few leucocytes but almost free from red cells. These capillary plugs completely obstruct the flow of blood through a large proportion of the glomeruli. The endothelial cells of the capillaries according to most pathologists show active proliferation, producing masses which project into the vascular lumina. The entire tuft is swollen and may almost fill Bowman's capsule. There is often an exudation of leucocytes into the interstitial tissue around the glomeruli. Bowman's capsule and the tubules contain masses of exudate and red cells, but the tubular epithelium itself shows little abnormality as a rule, beyond some slight cloudy swelling or fatty change.

Such a severe reaction must of course if wide spread cause a high degree of renal insufficiency. The following are among the chief features of the disease: retention in the blood of non-protein nitrogenous products (p. 8), edema, hypertension (ch. 16) with perhaps cardiac dilatation, acidosis, anemia, hematuria, albuminuria and casts (hyaline, blood, granular and epithelial), and the passage of a small quantity of urine with a high specific gravity. The majority of cases are much milder than this but all degrees of severity are encountered. Uremia occasionally terminates acute nephritis. In most instances however, recovery occurs, or the disease progresses to the subacute or chronic form.

CHRONIC GLOMERULO-NEPHRITIS

The histopathology varies according to the stage of the disease, but in general the picture is one of a healing or a healed inflammatory lesion of the glomeruli. Many of the latter show occlusion, due to the proliferation of endothelial cells, and hyaline

degeneration of some or of all their capillary loops, others are replaced by fibrous tissue, and yet others may show compensatory hypertrophy. Proliferation of the epithelium resulting in thickening of the membrane covering the glomerular capillaries, i.e., the visceral layer of Bowman's capsule, is a prominent feature. Similar changes are also seen in the parietal layer of the capsule, and the cavity may become obliterated. Sooner or later the tubules are the seat of hyaline or fatty changes as a result of the ischemia caused by the obliteration of glomerular capillaries. The lumina of the tubules become dilated and the tubular epithelium flattened. Some tubules may show attempts at regeneration, masses of newly formed epithelial cells appearing here and there. Glomerular regeneration is never seen. Sclerosis of arterioles and hypertrophy of their muscular coats are found in the later stages and are probably the result of the sustained hypertension which is such a common feature of the disease. Some vessels may show obliteration of their lumina as a result of active proliferation of their lining cells (endarteritis obliterans). Following is a summary of the characteristic clinical features of chronic glomerulo-nephritis. Compare with lists on pages 469 and 470.

(a) *Albuminuria*, usually not excessive (from 2 to 10 grams per day). Albumin globulin ratio (in urine) between 3 and 10 (p. 5). Epithelial, blood, hyaline and granular *casts*.

(b) *Renal insufficiency* (see p. 471), reduced ammonia production and consequent fall in the ammonia acid ratio, *acidosis*.

(c) *Edema* is often absent, when present it is not so prominent a feature as in nephrosis. Reduction of plasma protein and chloride retention are also less pronounced than in nephrosis.

(d) *Hypertension*, *cardiac hypertrophy* and *retinal changes* (see ch. 16).

(e) Large urine volume (*polyuria*) in earlier stages, *nycturia*, reduced volume (*oliguria*) in later stages, urine of low specific gravity, tending to become fixed around that of deproteinized plasma (1.009 to 1.011). There may be *glycosuria*.

(f) *Anemia*.

(g) *Uremia* usual cause of death.

Though the disease under discussion is primarily a glomerulitis, it will be recalled that the blood supply of the tubules is dependent solely upon the glomerular capillaries. In the earlier stages of the disease only a proportion of the capillary loops in a given tuft may be involved in the inflammatory process. The total available filtering surface will

be reduced thereby, but the tubules may still be adequately supplied with blood through previous capillaries. Later, when the glomerular flow becomes greatly curtailed, the tubules of that nephron undergo degeneration and atrophy. To this has been attributed the impairment of the concentrating function of the kidney (reabsorption of water) in glomerulo-nephritis. The reduced ammonia production, as well as the glycosuria and the fall in the serum calcium which sometimes occur in the disease can also be attributed to tubular involvement.

In some instances, chronic glomerulo-nephritis during its earlier stages shows signs and symptoms almost identical with those of chronic lipid nephrosis (p. 469), nitrogen retention, hypertension, etc., not appearing until later in the course of the disease. It is then referred to as the *nephrotic type* or *stage of glomerulo-nephritis*. Edema is a prominent feature. In such cases, apparently, the action of the toxin at first has been upon the tubules which undergo degenerative changes, the effect upon the glomeruli at this time being relatively mild.

THE NEPHROSES

(*Synonyms: tubular nephritis, degenerative Bright's disease, parenchymatous nephritis*)

This term embraces those several diseases of the kidney, acute, subacute or chronic, mild or severe, in which the predominant lesion is a *degenerative* one, implicating the *tubular epithelium*. Whereas the toxin of scarlet fever, for instance, acts primarily as a glomerular poison, several other toxic substances act chiefly upon the cells of the tubules. In many fevers albuminous degeneration ("cloudy swelling") of the tubular epithelium occurs, and albumin appears in the urine but disappears after the fever has subsided. This is referred to as *febrile nephrosis*. Similar degenerative changes of varying degrees of intensity occur in *diabetes*, *malaria*, *pernicious anemia*, *pregnancy*, *obstructive jaundice*, *intestinal obstruction*, etc. These together with the febrile nephroses, since they give rise to few or no symptoms, are often grouped under the term *larval nephroses* (L. larva, a mask).

A much more severe form of nephrosis (called *lower nephron nephrosis*) is caused by various poisons, e.g., salts of mercury, bismuth or uranium, carbolic acid, etc., and as a result of severe damage to muscle (e.g., the crush syndrome), burns, transfusion accidents, or peripheral circulatory failure. Extensive degeneration or necrosis of the tubular

epithelium occurs. In other instances the renal tissue undergoes amyloid degeneration. This *amyloid nephrosis* is seen in pulmonary tuberculosis, chronic pyogenic infections (e.g., osteomyelitis), the tubules show amyloid degeneration and atrophy. Many of the glomerular tufts may also be obliterated by amyloid deposits. The renal arterioles may show similar changes. Not only in amyloid nephrosis, but in some of the other more severe forms, e.g., those due to chemical poisons, the glomeruli may be involved in the degenerative changes, though usually to a much less degree than the tubules.

The clinical manifestations of the different types of nephrosis vary widely. Many are of mild degree and of a temporary nature, a small amount of albumin and a few casts are present in the urine and renal insufficiency is absent. The more severe and chronic forms are as a class characterized by *edema, a high degree of albuminuria, reduction of plasma protein* (mainly of the albumin fraction) *lipemia* and *hypercholesterolemia* and *the absence of hypertension and retinal changes*. Except in very severe types, such as those caused by chemical poisons in which necrosis of the tubular epithelium may occur accompanied by anuria, *renal insufficiency* is absent or of moderate degree.

LIPOID NEPHROSIS

(*Synonym: genuine or pure nephrosis*)

This is an exceedingly rare disease which occurs in children and young adults, did it not possess certain features of great clinical and physiological interest its description in a book of this nature would scarcely be justified. The histopathology is one of a pure degenerative disease of the tubular epithelium. The glomeruli are usually normal in appearance,² as are also the arterioles and larger vessels. The degenerative change consists of the deposition of an anisotropic lipid material (cholesterol esters) in the tubular cells and interstitial tissue. Cloudy swelling, necrosis or calcification of the epithelium are less usual findings. The chief clinical features are:

(1) *Edema*, frequently massive with *ascites*, protein content of edema fluid low (0.1 per cent). Low plasma

² Though generally looked upon as a distinct and separate disease, Bell and some other pathologists consider lipid nephrosis to be not clearly defined from the nephrotic type of glomerulonephritis, and that some glomerular damage always exists, as evidenced by the leakage of albumin through the glomerular membrane. Bell states that thickening and porosity of the basement membrane of the capillary loops can be demonstrated by appropriate staining.

albumin (less than 2.5 per cent) and total protein (less than 5 per cent).

(2) *Albuminuria*, often excessive (5 to 20 grams or more daily), albumin-globulin ratio in urine high (10 to 20) and low in plasma (see p. 5).

(3) *Urine* of small volume, of normal or high specific gravity and low in chlorides when edema is present. During recession of the edema the urine volume increases, the specific gravity falls and the chloride excretion rises. Hematuria is absent.

(4) Absence of hypertension, cardiac hypertrophy or retinal changes.

(5) *Lipemia* and *hypercholesterolemia* (up to 1000 mg or more per 100 cc), lipid droplets in the urine.

(6) *Low metabolic rate*.

(7) *Absence of renal insufficiency*, uremia does not occur, death usually results from some intercurrent infection, especially peritonitis (pneumococcal).

The cause of the disease is unknown. Two views are held: (a) that it is primarily a degenerative disease of the renal tubules resulting from an unknown toxin, (b) that it is a general metabolic disease of which renal manifestations are merely a part. Epstein believes that the protein in the urine differs from the normal plasma protein, being excreted as any "foreign" protein would be. But none of these claims has been substantiated. The disease may have its origin in a disorder of some endocrine function.

ARTERIOSCLEROTIC KIDNEY DISEASE

(*Synonym: nephrosclerosis*)

This occurs in three forms. (1) In elderly persons the *larger renal arteries*—arcuate, interlobular or interlobar—are chiefly involved in sclerotic changes as part of a general systemic arteriosclerosis. As a result of closure of some of these branches the blood supply may be shut off from relatively large renal areas (infarction). The nephrons within these areas atrophy and the reserve of the kidney becomes reduced. This is a picture of the *senile kidney*. The changes as a rule give rise to nothing more than a mild degree of renal insufficiency. The condition is of no great clinical or physiological interest.

(2) The *renal arterioles* become sclerosed (proliferation of their endothelial cells and a thickening of their muscular coats) as a sequel to *essential hypertension*. The persistent high pressure is looked upon as the direct cause of the arteriolar changes. The closure of numbers of glomerular arterioles causes a wide-spread destruction of glomerular tufts which become hyalinized or replaced by fibrous tissue. Degenerative changes occur in the tubules and the interstitial tissue is increased, but effects of an inflammatory nature are absent. Renal insufficiency results. When renal symptoms supervene

they resemble those of chronic glomerulo-nephritis, but tend to be overshadowed by the primary cardiovascular features. The following are the main differences between this type and chronic glomerulo nephritis

(a) Renal insufficiency is not, as a rule, a prominent feature until congestive heart failure supervenes

(b) Albuminuria slight, absent or intermittent.

(c) Edema is not as a rule a very pronounced feature and usually of cardiac origin

(d) Greater degree of hypertension and cardiac hypertrophy

(e) Death usually due to cardiac failure or cerebral hemorrhage, rarely to uremia

(3) As a result of *malignant hypertension* (p 166) the renal arteriolar disease is more intense. Endarteritis and necrosis of the afferent vessels (*malignant nephrosclerosis*) cause a more rapidly progressive destruction of the glomerular tufts. Renal insufficiency develops acutely, when death frequently occurs in uremia

Some of the more useful tests of renal function

A. TESTS INVOLVING AN EXAMINATION OF THE BLOOD
The non protein nitrogen (NPN) may be estimated in whole blood or in plasma, or the urea alone may be determined. The low mean and maximum values of the blood urea in health are, respectively, 10, 30 and 48 mg per 100 cc. of blood. In renal insufficiency it may be anywhere, according to the degree of impairment, from 60 to 500 mg per 100 cc. The severity of the insufficiency, whether relative or absolute, may be gauged by making the determinations while the patient is upon an ordinary diet and again when the intake of protein is reduced to the minimum. Examination of the blood will not reveal the earlier stages of renal insufficiency. It was found by Van Slyke and his associates that renal function may be reduced to 40 per cent or even to 25 per cent of the normal before a rise in the blood urea above the normal maximum occurs. The creatinine content of the blood (normal 1 mg per 100 cc.) is also sometimes employed as an index of renal function, but its rise occurs even later than that of urea. In severe degrees of renal insufficiency the creatinine may rise as high as 60 mg or more per cent, 5 mg is a more usual figure.

B. TESTS INVOLVING AN EXAMINATION OF THE URINE.

(a) Urea concentration test of Maclean and de Wesselow consists of giving 15 grams of urea in 100 cc. of water after the bladder has been emptied. The bladder is again emptied 1 hour and 2 hours later, the first urine specimen is discarded, since the urea induces a diuresis which lowers its concentration and so gives a false result. The degree of insufficiency is gauged from the percentage of urea in the second specimen. A concentration of 2 per cent or more indicates normal function. (b) Phenolsulphonephthalein is injected intravenously. In a normal person it dyes the urine in about 8 minutes, about 65 per cent or more being excreted within 2 hours

Its elimination is delayed when renal insufficiency exists. (c) *Specific gravity test*. This is one of the oldest and at the same time one of the simplest tests of renal function (see pp 464 and 472)

The dilution and concentration tests of Volhard and Strauss. The *water or dilution test* consists in determining the volume and specific gravity of the urine each hour for 4 consecutive hours after the ingestion of 1500 cc. of water. A normal person excretes the entire 1500 cc. within 2 or 3, or at the longest 4 hours, and most of it within the first hour. The specific gravity of the urine is from 1.001 to 1.003. The patient with renal insufficiency takes a longer time to eliminate the extra water, and the specific gravity of the urine is higher (p 472). The *concentration test* is usually performed after the dilution test and consists in determining the specific gravity of the urine after fluids have been withheld for 10 hours or more. The normal person concentrates the urine to a specific gravity of 1.025 to 1.032. A specific gravity below 1.020 indicates impaired concentrating power—defective reabsorption by the tubules.

Urea clearance test. This was introduced by Austin, Stillman and Van Slyke as a measure of renal function. The quantity of blood "cleared" of urea per minute is referred to as the *blood urea clearance*. By the term clearance is meant the quantity of blood which would give up all its urea or other substance which it contained did such a volume of blood conceivably pass alone through the kidney each minute. Of course, the total volume of blood flowing through the kidney is many times greater (around 1200 cc per minute) than the quantity specified and gives up only a small fraction of its urea each minute. For example, if the blood urea concentration is 40 mg per 100 cc and a total of 30 mg of urea pass into the urine per minute, it can then be said that the quantity of urea removed from the circulation per minute is equivalent to that contained in 75 cc of blood.

It has been shown that above a certain magnitude of urine flow, called the *augmentation limit*, urea excretion is maximum, urea excretion is unaffected by a further increase in urine flow. The quantity of blood cleared of urea remains constant. Thus, the *maximum clearance* (Cm) is calculated from the following

$$UV/B = \text{blood urea clearance (quantity of blood cleared of urea N per minute)}$$

where U is urea N in mg per 100 cc. of urine, V, the urine volume in cc., and B, mg of urea N per 100 cc of blood. Thus, if the urine volume is 3 cc per minute and the urea N per 100 cc of urine, and of the blood, 330 mg and 13.2 mg respectively, then

$$(330 \times 3)/13.2 = 75 \text{ cc maximum clearance (Cm) per minute}$$

With lower urine volumes (0.5 to 2 cc. per minute) the quantity of urea excreted becomes progressively re-

duced due to the back diffusion of urea across the tubular membrane into the blood. The excretion down to a value of 0.35 cc per minute is proportional to the square root of the urine volume (V). Thus, with an ordinary urine flow (1 cc per minute) the urea clearance—the so-called *standard clearance* (C_s)—is calculated from the formula

$$(U \sqrt{V})/B = \text{standard clearance } (C_s)$$

The standard clearance averages normally 54 cc and the maximum clearance 75 cc per minute.

The urea clearance test is of no value in detecting renal insufficiency in its earlier stages, for no reduction in the clearance occurs until less than 50 per cent of functioning renal tissue remains intact.

In advanced nephritis the urea clearance at a urine flow of 1 cc per minute may be as low as 9 cc, and the maximum clearance (urine volume above 2 cc per minute) only from 10 to 12 cc., or from 10 to 15 per cent of normal.

Inulin clearance This was introduced by Homer Smith and his colleagues to measure the rate of glomerular filtration. Inulin, a starch-like polysaccharide with a relatively small molecule (mol wt 5100) passes freely across the glomerular membrane. It is not metabolized, is neither excreted nor reabsorbed by the tubules, and does not combine with the plasma proteins. Its clearance amounts to around 130 cc. per minute, and under all conditions (e.g. rate of urine flow or plasma concentration) its clearance is equal to the rate of filtration. Mannitol may be used for the same purpose. Thus, either of these agents serve as an invaluable means of determining the filtration rate or of measuring the reabsorption or tubular excretion of a given urinary constituent. For example, a substance, such as creatinine, diodrast, or phenol red, which has a clearance *greater* than that of inulin, must be, in part at least, excreted by the tubules, if the clearance of a substance such as urea or glucose is *less* than that of inulin, it must have undergone partial or complete reabsorption. Inulin is given by intravenous injection.

Tubular transfer tests When the plasma level of glucose is raised progressively, a point is reached where no more can be reabsorbed by the tubules, glucose then appears in the urine. When the quantity of sugar reabsorbed remains constant with rising plasma concentration the kidney has become saturated, i.e., the maximal capacity of the tubules to reabsorb glucose has been reached.² This maximum, which is about 325 mg per minute for the human kidneys is called the glucose T_m (maximal transfer) or T_{m0} . The actual quantity of glucose reabsorbed per minute in mg (T_G)

is the difference between the quantity filtered in mg per minute filtered load and that excreted in mg per minute in the urine (U_0V , where U_0 is the concentration of glucose in the urine and V , the urine volume in cc. per minute). The quantity of glucose filtered per minute is the product of its concentration in the filtrate⁴ in mg per cc. (P_G) and the rate of glomerular filtration in cc per minute as determined from the inulin clearance (C_{in}). Thus,

$$T_G = P_G C_{in} - U_0 V$$

Up to the point of maximum reabsorption T_G equals approximately $P_G C_{in}$. In other words, under ordinary circumstances, all the glucose of the filtrate is reabsorbed by the tubules.

In a corresponding manner, the rate of the tubular excretion of diodrast can be ascertained. It will be the rate of the excretion of diodrast in the urine ($U_D V$), less the rate of diodrast filtration. This latter is the plasma concentration (P_D) multiplied by the rate of glomerular filtration (C_{in}) by the fraction of free *filterable* diodrast in the plasma (FW). Thus,

$$T_D = U_D V - FWP_D C_{in}$$

The maximum reabsorption of glucose (glucose T_m) is a measure of the reabsorptive capacity of the kidney, *the tubular absorptive mass as it is termed*. The diodrast T_m is used in a manner corresponding to the use of glucose T_m to estimate the mass of functioning excretory tubular tissue—the *tubular excretory mass*.

THE PATHOLOGICAL PHYSIOLOGY OF SOME OF THE MAIN FEATURES OF RENAL DISEASE

I RENAL INSUFFICIENCY

Renal insufficiency means a reduced capacity of the kidney to carry out its functions, namely

(a) *The excretion of non-protein nitrogenous substances*—urea, uric acid, creatinine, etc. The rate of excretion of urea is directly proportional to (i) its concentration in the blood when the urine volume is above a certain amount, called the *augmentation limit*, (ii) the square root of the urine volume when this is below the augmentation limit and above a urine flow of about 0.35 cc. per minute, when a sharper reduction in urea clearance occurs, it then falls in direct proportion to the urine flow, rather than to the square root of the

² Glycosuria of moderate degree occurs a little before maximum reabsorption has been reached. This is attributed to a certain imperfection in the balance of glomeruli as to the filtering capacity of tubules (see glomerulotubular balance, ch 35).

⁴ The concentration in the plasma is taken as the concentration in the filtrate. The latter is higher than the plasma concentration owing to the removal of the plasma proteins during filtration and a small correction (W) is sometimes required. The value of W is 1 per cent of plasma protein/100. This correction is always necessary in the calculation of the excretion of diodrast, which is, in part, bound to the plasma proteins.

latter, and (iii) the amount of active renal tissue when this is less than about 50 per cent of the normal. The augmentation limit in man *normally* averages about 2 cc. per minute. At or above the augmentation limit the urea excretion in any individual is at a maximum, being unaffected by further increase in urine flow.

(b) *The excretion of water* Though there may be polyuria (p 477) which might suggest that there is no incapacity of the kidney to eliminate water, it will be found that the inefficient kidney cannot respond promptly when an extra demand is made upon this function. Consequently, when a large quantity of water is drunk it is not excreted within the normal time. Furthermore, the specific gravity of the urine does not fall to the low level (1.001 to 1.003) of urine formed in health under the same circumstances, but remains around that of deproteinized plasma (glomerular filtrate), namely, 1.009 to 1.010.

(c) *Concentrating power* (reabsorption of water) is impaired. The low threshold substances, such as urea, uric acid, etc., are in lower concentration in nephritic than in normal urine even when the concentrating power of the kidney is encouraged by reducing the fluid intake. If fluids are withheld for some time from a healthy person, he passes urine of a high specific gravity, 1.030 or over. The severely impaired kidney is unable to raise the specific gravity much, if at all, above that of protein-free plasma. Taking the two last mentioned functions, (b) and (c), together it may be said that as the renal injury progresses, the specific gravity of the urine after the ingestion of a quantity of water tends to be depressed less and less toward the normal figure of about 1.002, when fluid is withheld the specific gravity is raised with increasing difficulty toward the normal maximum of about 1.030. In other words, instead of showing the great swings in specific gravity of normal urine under varying conditions of fluid intake, nephritic urine tends to become fixed at a specific gravity around 1.010. *Hyposthenuria* is the term used to denote subnormal concentrating power. *Isothenuria* is a term which means that the kidney cannot form urine with a higher or lower specific gravity than that of protein-free plasma.

Normally, the tubular fluid becomes hypertonic by the reabsorption of water from the nephron distal to the proximal convoluted tubule. It is impairment of this part of the nephron that is responsible for the failure in concentrating power of the kidney. But the insufficient kidney is also un-

able to *reduce* the specific gravity of the urine much below that of the filtrate, which would suggest that the ability of the tubules to increase the reabsorption of chloride and other salts in response to the ingestion of water is also impaired. The urine in the later stages is nearly isotonic with the plasma under all conditions, the reabsorbed fluid, therefore, has also a concentration in solids similar to that of deproteinized plasma under all circumstances. The tubule cells have largely lost their power of selective reabsorption, substances in the filtrate being returned to the blood in the peritubular capillaries by a simple process of diffusion.

The first demonstrable impairment of renal function in *glomerulo nephritis* is a reduction in the filtration rate, which occurs before the concentrating power of the tubules is affected. The lowered filtration rate is associated with a reduction in the filtration fraction (p 454), which may be no more than 8 or 9 per cent, i.e., less than half of normal. Since the renal blood flow is normal or increased at this stage, the subnormal filtration rate is not the result of lowered glomerular pressure (as would result from afferent arteriolar constriction), but to structural changes in the filtering bed (obliteration of capillaries and thickening of the glomerular membranes). The reduction in filtration rate (glomerular damage) at this stage while tubular function is little impaired leads to the retention of water and electrolytes, the urine volume is then below normal. The ratio, glomerular filtration/diodrast Tm (GF/Tm) is depressed. As the disease advances, glomerular and tubular damage run more or less parallel, and the GF/Tm ratio remains low. Later, tubular damage becomes more pronounced, and the GF/Tm ratio increases, and if arterial hypertension supervenes, the glomerular pressure rises and the blood flow diminishes as a result, apparently, of efferent arteriolar constriction, the filtration fraction increases and the GF/Tm ratio is elevated to normal or above. In these later stages of chronic glomerulonephritis, as a result of the deterioration of tubular reabsorptive function, urine volume rises and there is a loss of body water and electrolytes.

The disproportionate damage to the glomerulus and tubules in glomerulonephritis causes varying degrees of functional imbalance between these two parts of the nephron (see p 450), and such imbalance accounts to a large extent for the disturbances in renal functions which are observed. Severely injured glomeruli, for example, filtering into relatively normal tubules will result in those

nephrons producing a small volume of urine of normal composition. Conversely, if the tubules are the more severely damaged parts, the filtrate in passing along the tubules will be little altered, the urine formed by such nephrons will be of relatively large volume and resemble the filtrate in composition.

The nephrotic stage of glomerulonephritis in adults shows disturbances similar in renal function to those described above. But in children with nephrosis the urea and inulin clearances are actually greater than normal. The renal blood flow and filtration fraction are both increased.

In renal disease due to hypertension (*nephrosclerosis*) the effects upon renal processes contrast sharply with those found in chronic glomerulonephritis. Tubular insufficiency precedes any pronounced reduction in filtration rate. The earliest change is a fall in renal blood flow and an increase in the filtration fraction (p 454)—indications of efferent arteriolar constriction. Later, as a result of ischemia, tubular damage becomes evident, reabsorption is subnormal and, as a consequence, polyuria occurs. The constriction of the efferent arteriole is not affected by sympathectomy, but is abolished by the administration of a pyrogen, the renal blood flow increases. These facts strongly suggest the action of a pressor substance, probably renin.

When one speaks of renal insufficiency it is the impairment in the functions referred to in the foregoing paragraphs that is meant, since it is for the measurement of the degree of impairment of these functions that the most reliable methods have been devised. Yet, it should be remembered that renal disease does not single out any particular function for attack, when one is impaired others also suffer more or less. *Nitrogen excretion, water elimination, concentrating power, ammonia and hippuric acid production, and base conservation* (p 460) are all interfered with more or less together. There is one apparent exception, the leakage into the urine of serum protein—proteinuria—which is an indication of some abnormality of the glomerular membrane, may be present in the absence of other signs of disturbed renal function (see *lipoid nephrosis* and *benign albuminuria*). Proteinuria, therefore, though a fault in renal function is not a criterion of renal insufficiency in the ordinarily accepted sense of this term.

Renal insufficiency may be *relative*, as when there is an accumulation of nitrogenous products in the blood, and water elimination is difficult upon

an ordinary diet, but the kidney performs its work if the water intake and the protein of the diet are reduced to the minimum. When the insufficiency of the kidney is evident upon a rigidly restricted diet, it is said to be *absolute*.

II ALBUMINURIA (OR MORE CORRECTLY PROTEINURIA)

This is one of the commonest accompaniments of renal disease, but it may also occur in healthy persons.

Albuminuria in kidney disease

The urinary protein is derived from the plasma and there is abundant evidence that it passes mainly or entirely through the glomerulus rather than through the tubular epithelium, though in severe tubular damage a part of the urinary protein is that which would otherwise have been reabsorbed. The quantity of albumin in the urine usually exceeds that of globulin, for albumin having a much smaller molecule escapes more easily through the "pores" of the glomerular filter. The still larger molecule of fibrinogen rarely escapes into the urine, and when fibrin is found in the urine it is usually due to hemorrhage into the urinary tract. The number of milligrams of albumin in a liter of urine divided by the globulin content in milligrams gives what is known as the urinary albumin-globulin (A/G) ratio. This has a value of between 3 and 10 in chronic nephritis, but is from 10 to 20 in chronic (lipoid) nephrosis. In amyloid nephrosis it is usually quite low—between 1 and 2. A ratio below 5 in chronic nephritis appears to be a grave prognostic sign. When the loss of protein in the urine becomes excessive, as in lipoid nephrosis and in the nephrotic type of nephritis, the serum albumin cannot be regenerated sufficiently rapidly to maintain its normal level in the plasma, a deficit therefore occurs (see p 475). The albumin-globulin ratio of the *plasma* falls as that of the *urine* rises. In the plasma the concentration of the globulin fraction may actually exceed that of albumin fraction, i.e., the ratio is inverted and may be as low as 0.2. The proteinuria of lipoid nephrosis is difficult to explain, since the glomeruli usually appear normal, whereas the tubules show marked degenerative changes. It is assumed, nevertheless, that some ultramicroscopic change in the glomerular membrane has occurred which is responsible for its abnormal permeability. A certain small portion of the protein which escapes

into the filtrate is reabsorbed, unless, as mentioned above, the tubules are seriously damaged. The reabsorbed protein, or products of its intracellular digestion, are seen as hyaline granules or droplets, which are responsible, according to Oliver, for the "cloudy swelling" of the tubule cells, the name given to a relatively mild nephrotic lesion long described by pathologists (ch 35)

Albuminuria in the absence of kidney disease

This form of albuminuria has been distinguished from the preceding by the use of various qualifying terms—*adolescent, postural, orthostatic, benign, functional, physiological*, etc.—which collectively denote its chief features. The albuminuria should be looked upon as being due to some minor disorder of renal physiology rather than the result of any definite structural change in the kidney. It may occur at any age, but the subjects are usually young—between the ages of 8 and 18—and in good health, there being no evidence of renal damage. It is most likely to occur after exercise. The condition does not lead to kidney disease. MacLean found the condition in 4 per cent of healthy British soldiers. It was found in 5.25 per cent of 16,748 healthy freshmen examined at the University of Minnesota (Diehl and McKinley). The protein is usually small in amount (0.2 per cent), and is commonly present only when the subject is in the erect posture. The condition is probably due to some disturbance in the renal circulation leading to congestion (venous stasis) induced by the upright position, with or without *lordosis*. The latter is considered by some as an important factor. That the albuminuria is of circulatory origin is indicated by the following facts: (1) compression of the renal vein in animals or the injection of adrenaline which constricts the renal vessels, causes albumin to appear in the urine, (2) Erlanger and Hooker observed that the pulse pressure was lower than normal in subjects of orthostatic albuminuria and that the degree of albuminuria increased in proportion to the reduction in pulse pressure. The erect position lowers the pulse pressure (p 155). In benign albuminuria, the albumin is little if at all in excess of the globulin, i.e., the urinary A/G ratio is low, between 1 and 2 or less. Almost any normal person may pass protein in the urine (as shown by the ordinary clinical tests) after severe muscular exertion or cold bathing. Sometimes the protein of the urine is derived from ingested protein, as after eating a large number

of raw eggs. In animals protein appears in the urine after the intravenous injection of egg white (the molecular weight of egg albumin is relatively small, about 35,000). The benign albuminuria of pregnancy is probably due to interference with the renal venous return. Albuminuria occurs in congestive heart failure as a result probably of slowing of the renal blood flow and a certain degree of ischemia. It also occurs in anemia, fevers, etc. and in these instances there is frequently cloudy swelling of the cells of the tubules, the renal condition constituting a mild type of nephrosis (pp 468 and 469).

As mentioned on page 450, most normal urines contain traces of protein which require, as a rule, especially delicate methods for their detection.

Bence-Jones proteins. This protein is not normally present in the plasma except perhaps in traces but appears in the urine in cases of certain bone tumors (myelomas). The protein has a small molecule (molecular weight around 35,000) and escapes readily through the healthy glomerular membrane. When urine containing it is acidified with acetic acid and heated to 60°C., flocculent coagulum appears but disappears again when the temperature is raised to the boiling point. When the urine is allowed to cool, cloudiness reappears but again dissolves as the temperature falls to around 40°C.

III EDEMA IN KIDNEY DISEASE

The general physiological principles involved in the passage of fluid through the capillary membrane and the factors concerned in the production of edema have been dealt with in chapter 3. *Without doubt the outstanding factor in the production of the edema of chronic renal disease is the depletion of the serum proteins* which results from the excessive proteinuria. The loss of albumin, since it has a smaller molecule and so exerts a greater osmotic pressure, is a more important factor in the production of edema than is the loss of globulin.

The osmotic pressure exerted by 1 per cent of the combined proteins of normal plasma is around 4 cm H₂O. But the osmotic pressure of a 1 per cent solution of serum albumin is 7.5 cm H₂O, whereas the osmotic pressure exerted by the same concentration of the globulin fraction is only 2 cm H₂O, and that of fibrinogen, owing to the large size of its molecule, is still less. The albumin fraction exerts over 80 per cent of the total oncotic pressure of normal plasma. These figures, together with the fact that serum albumin is diminished to

a greater degree than is the serum globulin, explains why a fall in the total protein of the plasma in nephrosis produces a much greater effect upon the oncotic pressure than one would be led to expect from a simple calculation based upon variations in the total protein concentrations of normal plasma. Furthermore, even in normal plasma E. H. Fishberg has found that, as the total protein concentration decreases in arithmetical progression, the oncotic pressure diminishes in geometric progression.⁴

It is just those types of chronic kidney disease (nephrosis and the nephrotic type of glomerulonephritis) in which large quantities of albumin are lost in the urine and the plasma protein is low, that show the highest degrees of edema. In other types of chronic glomerulonephritis and in arteriosclerotic kidney disease there is little tendency to severe serum protein deficit, and edema, attributable to the renal lesion, is unusual. A fall, even though pronounced, in the serum albumin is not always followed by edema, since the fibrinogen and globulin fractions may show compensatory increases and raise the total protein of the plasma sufficiently to prevent a fall in the plasma osmotic pressure below the critical level of 20 mm Hg (normal 25 to 30 mm). But it is generally conceded that a reduction in the albumin fraction below 2.5 per cent (normal 4 to 5 per cent) and a total protein content of around 5 per cent (normal 7 to 8 per cent) is the point at which edema commences. Krogh cites a case of nephrosis with edema in which the urine contained nearly 3 per cent of protein, the total plasma protein was 5 per cent, giving an osmotic pressure below 8 mm Hg. The capillary blood pressure was considerably higher than the osmotic pressure, namely, 11 mm Hg. According to Peters, when the total protein falls below 4 per cent the edema is intractable to all forms of treatment. Moore and Van Slyke have shown that a linear relationship exists between the specific gravity of the plasma and its total protein content, the critical specific gravity for edema formation is 1.023 (normal average 1.027). The edema fluid of chronic renal disease is very low in protein (0.1 per cent or less) and contains about the same sodium chloride concentration as does

plasma. During the discharge of the edema fluid the excretion of sodium in relation to that of potassium rises, an indication that the excess fluid is of extracellular origin (p. 20).

Edema has been produced experimentally in dogs by placing them upon a low protein diet (p. 39) or by plasmapheresis (p. 7). The edema produced by the latter procedure disappears when, as a result of regeneration of the plasma proteins, the osmotic pressure rises above the critical level, the nutritional edema is abolished after the animals have been given an adequate supply of protein. It has also been demonstrated that a low protein diet may precipitate edema in a nephrotic patient who had previously been free from it. Conversely, a high protein diet tends to raise the concentration of plasma proteins in nephrotic subjects and of causing a reduction in the edema. A high protein diet which will encourage the regeneration of plasma protein would appear, therefore, to be a rational procedure in the treatment of nephrotic edema. It also follows that it is undesirable to remove large quantities of edema or ascitic fluid mechanically (paracentesis) for, though the protein content of the fluid is very low, the total amount lost to the body through such a procedure may be significant.

But the formation of edema in renal disease is a complex problem, and it must be admitted that it can not always be explained upon the basis of a low serum protein, though such is probably always present in some degree when the total plasma protein is less than 5 per cent, it occurs in some cases with a protein concentration of over 7 per cent.

The distribution of the edema in renal disease differs from that of cardiac failure. In the former the fluid tends to collect in sites where the tissues are loose and inelastic, e.g., beneath the eyes, and is commonly more pronounced after the night's sleep than when the patient is up, drainage of the fluid then being aided by gravity. In cardiac edema gravity and raised capillary pressure encourage the seepage of fluid into dependent parts—ankles and legs—and it is therefore more pronounced when the patient is upright, and is reduced or disappears in recumbency.

The relation of salt to edema

It will be recalled that water can be retained in the body only as an isotonic solution, and the administration of sodium chloride to the patient

⁴This relationship is expressed in the following equation

$$p = \frac{c}{v}, \quad \text{where}$$

p is the reciprocal of the osmotic pressure, v the reciprocal of the protein concentration, and a a constant dependent upon the particular serum.

on the verge of edema precipitates the condition, the edema disappears again upon a salt-free diet.

Sodium bicarbonate has an effect upon edema formation similar to that of sodium chloride, whereas other chlorides (e.g., KCl, NH_4Cl) either have no water retaining influence or through their acidifying effect (pp 24 and 455) cause diuresis and the discharge of the edema fluid. It has also been shown by Hastings and Van Dyke that edema can be produced in dogs by the administration of sodium bromide. These and other facts have shown that the influence upon the production of edema is exerted by the sodium rather than by the chloride ion.

Edema can be produced even in healthy persons by the administration of very large quantities of salt followed by the drinking of an abundance of water. Baird and Haldane took sodium chloride (38 grams) or sodium chloride and bicarbonate in 500 cc of water. If, while the diuresis which followed the salt ingestion was subsiding, 2½ liters of water were drunk, there was no apparent effect, the urine volume continuing to fall. Edema about the ankles was noted. When, however, an additional 500 cc were drunk, profuse diuresis resulted. It was also found that water or isotonic saline caused a much greater diuresis than hypertonic saline. They give as the probable explanation of their results that the salt was first deposited as a hypertonic solution in the tissues, the water which was ingested later left the capillaries to produce an isotonic tissue fluid. Water ingested in excess of this requirement was excreted. The smaller diuretic effect, i.e., the greater water retention, of hypertonic saline as compared with that induced by water or hypotonic saline finds a similar explanation. (See also water diuresis, ch 35.)

It is seen, therefore, that even in health the balance between intra- and extra-capillary fluid can be upset and a temporary edema produced. Normally fluid is forced out of the capillary near the arterial end where the blood pressure overbalances the colloid osmotic pressure (ch 3). At the venous end, where the blood pressure no longer exceeds the pressure of the tissue fluid and the osmotic pressure of the plasma proteins, water is taken up. The fluid poured into the tissues as a result of the deposition of chloride, therefore, soon undergoes reabsorption. When, however, the concentration of plasma protein is low, reabsorption becomes more difficult. The water which has migrated from the vessels to dilute the tissue fluids to isotonicity is not readily removed. Deficit in

the plasma protein is therefore the primary cause of nephrotic edema, salt retention in the tissues is a contributory factor, salt, like water itself plays a passive rôle simply enabling more isotonic tissue fluid to be formed i.e., providing one of the essential constituents of the edema fluid. In other words, water can only be retained as an isotonic fluid, if, therefore, salt is withheld the excess water is excreted. It is also recognized that in certain dehydrated states ingested water is not retained unless salt is also supplied.

In nephrosis with massive edema, injections of a colloid solution with an osmotic pressure equal to that of normal plasma are employed, together with rigid salt restriction, the former to restore the plasma oncotic pressure and thus to hold fluid within the vessels, the latter to limit the quantity of extracellular fluid which can be formed. In the less severe forms of nephritic edema a more moderate curtailment of the salt intake may be combined with the ingestion of a carbolyte (e.g., amberlyte) or sulfonic ion-exchange resin. Ammonium ions held by the resin are exchanged in the intestinal tract for sodium ions derived from the food and digestive secretions, the Na^+ is excreted in the feces.

In *acute nephritis*, especially in the earlier stages, edema may occur in the absence of plasma protein deficit. The edema is not widespread as a rule, but may appear before the renal manifestations. It is considered as due to increased permeability of the capillary wall, the high protein content of the edema fluid (1 per cent) supports this conclusion. It is likely that the toxic substance which has produced the inflammatory lesions in the glomerular capillaries has also injured the systemic vessels. It is significant that poisons, such as the toxins of diphtheria and scarlet fever, and other substances which cause glomerular damage, are also general capillary poisons.

IV CASTS, CYLINDURIA

Part of the protein material which has passed through the damaged glomerulus undergoes coagulation and becomes moulded in the distal tubules and the collecting tubules to form the microscopical urinary bodies known as casts. Blood cells which have escaped from the capillaries of the glomerular tuft, epithelial cells shed from the tubules, or lipid material, may become moulded in a similar way. The several varieties of casts—*blood*, *epithelial*, *fatty* or *waxy*—are named according to the material of which they are composed. *Hyaline* casts are formed of coagulated albumin and are translucent and homogeneous. *Granular* casts are made of a

conglomerate of blood or epithelial cells in a matrix of albumin. In the terminal stages of renal failure the shed epithelial cells of the larger collecting tubules form exceptionally large elements termed *renal failure casts* (Addis). The urine may also contain many free red and white cells, the numbers varying according to the type and severity of the kidney lesion. Nevertheless, the presence in the urine of relatively small numbers of hyaline casts or even of a few blood cells is by no means an indication of renal disease. Addis examined the urine of 74 healthy medical students and found hyaline casts in 45. The figures for the rates of excretion for the entire group are given in table 35.

The protein excreted in the urine in renal disease is not ordinarily coagulated, special means, e.g., heat, nitric acid, etc., must be employed to bring this about. Why protein forms a coagulum in the tubules is a question of long standing but which has received a partial answer from the researches of Oliver. It was found that hyaline casts give the metachromatic reaction with toluidine blue (p. 116) which is characteristic of chondroitin-sulfuric and mucosin-sulfuric acids. These acids have the property of coagulating albumin. The casts are formed chiefly in the distal convoluted and collecting tubules, for in this situation not only is the protein more concentrated but also the reaction of the urine approaches the isoelectric point of albumin. These two factors alone, however, are not sufficient to cause precipitation. Chemical identification of the metachromatic substance, which is apparently the determining factor in the formation of the coagulum, has not been made, nor has its source been determined. Whether or not it is one or other of the mucopolysaccharides mentioned above remains to be shown.

V POLYURIA, OLIGURIA AND ANURIA

As stated on p. 472, the earliest demonstrable change in function of the glomerulonephritic kidney is a reduction in the filtration rate, tubular concentrating function remaining little, if at all, affected. This results in a reduction of urine volume. Later on in the disease the urine volume increases. An increase in the volume of the night urine—*nycturia*⁶—is often one of the first signs of a failure in concentrating power. The normal kidney concentrates the urine during sleep to a greater degree than during waking hours, so that the urine volume does not exceed the capacity of the bladder, but the nephritic kidney, as a result of tubular damage, finds this impossible. *Polyuria*, or the production of abnormally large amounts of dilute urine during the day and night, is characteristic of

TABLE 35

Rate of excretion of casts, blood and epithelial cells per twelve hour period
(After Addis)

	AVERAGE	LOWEST	HIGHEST
Casts	1,040	0	4,270
Red blood cells	65,750	0	425,000
White blood and epithelial cells	322,500	32,400	1,835,000

the later stages of glomerulonephritis. Unlike that of diabetes insipidus, the polyuria of chronic renal disease does not respond to pituitrin. In still later stages of renal insufficiency, or when edema is forming, the output of urine during the 24 hours is reduced considerably below the normal.

The polyuria of chronic nephritis suggests a compensatory response of the kidney to the reduction in the number of functioning renal units. Certain experimental observations seem to support this conception.

Rose Bradford found that when a part—sometimes as small as one-eighth or one-fourth—of the renal substance of an animal was excised, the remaining tissue responded within 24 hours by excreting large quantities of very dilute urine. The polyuria persisted throughout the period of observation, which extended over a period of from 1 to 2 years. When the amount of excised renal tissue was small, urea retention did not result. Verney and his associates carried out similar experiments upon isolated kidneys perfused by means of a heart-lung preparation. When a primary branch of the renal artery of one kidney was ligated, 50 per cent of its tissue was put out of action. Nevertheless, though the blood flow was reduced by one-half, there was no reduction in the quantity of urine formed by this kidney. It produced as much urine as the sound kidney of the opposite side. In other words, the active kidney tissue which remained immediately doubled its output and compensated fully for the functionless portion. The urine flow gradually increased throughout the experiment until finally it was some 30 times greater than that of the sound kidney. The composition of the urine from the injured kidney was altered. The urea percentage as well as the total quantity excreted showed a profound fall. It appears, however, that per unit of functioning tissue, the affected kidney excreted slightly more urea than the normal one. The chloride percentage was unchanged, so that the absolute amount, i.e., the amount per unit of time, excreted by the functioning half of the kidney was much greater (4 times in one experiment) than that excreted by the kidney before ligation, or than that excreted by the kidney of the opposite side. The specific

⁶ This applies, of course, only under ordinary conditions of fluid intake. A healthy person may have to get up in the night to pass urine if he has drunk much fluid before retiring.

gravity of the urine approached that of deproteinized plasma. In brief, the injured kidney excreted much more water and chloride and much less urea than the normal organ. Polyuria and hyposthenuria were also produced in dogs by Hayman and his colleagues by the surgical reduction of kidney substances; they found that a reduction in renal blood pressure or an increase in plasma colloids (thus reducing filtration) diminished the urine volume and raised the specific gravity of the urine.

The results of the experiments just cited are due apparently to intrarenal vascular changes (either of nervous or of humoral origin) through which glomerular pressure is raised and the volume of filtrate formed by the remaining glomeruli increased, in order to maintain the excretion of urea and other plasma constituents at the normal value. The greater amount of filtrate formed by each remaining glomerulus floods the corresponding tubules, and thus causes a more rapid tubular flow, with the result that reabsorption of water is reduced.

The polyuria of nephritis can be similarly explained. But, in contrast to the experiments just quoted, in glomerulonephritis the nephrons which are not completely destroyed by the disease are far from normal, added to the effect of a larger volume of filtrate being poured into a reduced number of tubules, and thus causing rapid flow along the tubular lumina, concentrating power is also defective as a result of the tubular damage itself. The large urinary flow is in a way compensatory and beneficial, the flooding of the tubules, for example, tending to diminish the reabsorption of urea by passive back diffusion, and thus to increase excretion, while the low osmotic pressure of the tubular fluid reduces the osmotic work which the tubule cells are called upon to do. But the excretion of a large volume of urine has its disadvantages, since it leads to dehydration and a serious loss of essential electrolytes.

In the polyuria of *nephrosclerosis* the filtration rate is normal, slightly reduced, or even increased, while the reabsorption mechanism is impaired as a result of tubular ischemic damage (constriction of efferent vessels). Owing to efferent arteriolar constriction and also to the high systemic blood pressure, the glomerular pressure is raised and the filtration fraction $\left(\frac{\text{filtration rate}}{\text{renal plasma flow}} \right)$ is increased (p 454).

Anuria. Complete failure of the kidney to form urine is called *suppression of urine*, or *anuria*. Anuria may occur in the final stages of chronic glomerulo-nephritis, the urine volume becoming gradually reduced to zero. Or it may supervene suddenly, as in acute nephritis, as a result of severe trauma (*traumatic anuria*), after surgical operation, following a transfusion of incompatible

blood, as part of the "crush syndrome" (p 305), or from poisoning, especially with bichloride of mercury. It occurs when the filtration rate is greatly reduced, and reabsorption of the small volume of tubular fluid is complete. After poisoning of the frog's kidney by mercury the filtrate is entirely reabsorbed. It is strongly suspected that traumatic anurias and those caused by the transfusion of incompatible blood, as well as the anuria of the "crush syndrome", or tourniquet shock, are due to ischemia of the renal cortex, as a result of the diversion of blood through the juxtamedullary glomeruli. That anuria may at times be of nervous origin has been suggested by many observations. The clear cut experiments of Trueta and his associates have shown that a tourniquet applied to a limb causes constriction of the femoral artery proximal to the tourniquet, of the artery of the opposite limb and of the aorta and renal artery. The narrowing of the vessels was observed by radiography after the injection of a radio-opaque material. The response of the renal artery was absent if the splanchnic nerves had been previously divided. Associated with the narrowing of the renal artery with its innervation intact, a diversion of blood from the more superficial glomeruli of the renal cortex to juxtamedullary glomeruli (p 442) was clearly demonstrated. When dye or India ink was injected, the medullary region became deeply stained, whereas the cortex remained pale and the time taken for injected radio-opaque material to appear in the renal vein was much shortened. Corresponding effects were obtained by stimulating the central end of the cut sciatic nerve (fig 361).

VI UREMIA

Two types of uremia are recognized—*true uremia* and *false* or *eclamptic uremia*.

True uremia is the terminal manifestation of renal failure. Its principal clinical features are anemia, dyspnea, anorexia, nausea and vomiting, weakness and twitchings of the muscles, occasionally convulsions, dizziness, mental disturbances, and finally, somnolence, stupor and coma with periodic breathing. The high urea concentration of the sweat causes the deposition of crystals of urea upon the skin—"urea frost"—which causes itching. The uremic state is preceded by anuria, or the passage of a small volume of urine of low specific gravity (around 1.010), a rise in the urea (usually to 150-500 mg per cent) and other non-protein

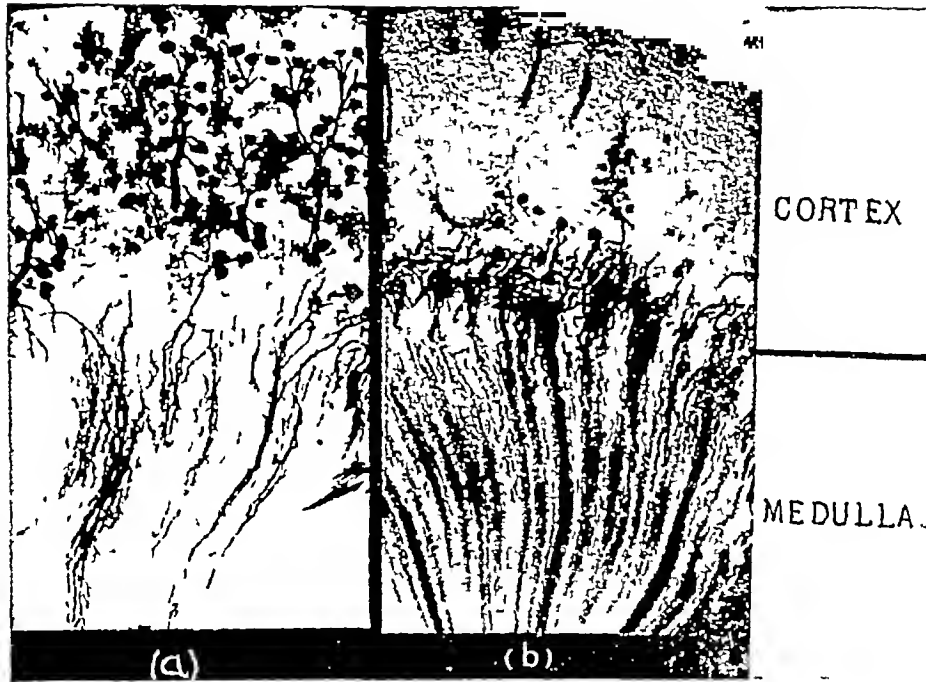


FIG 36.1 Showing (a) the distribution of India ink injected during life into the renal artery of a normal rabbit and (b) during the stimulation of the plexus of nerves surrounding the artery (After Trueta and associates, slightly modified)

nitrogenous compounds (e.g., creatinine and uric acid) of the plasma, as well as of phosphate occurs. The urea and creatinine clearances are greatly reduced.

Though the symptoms are associated with a high degree of nitrogen retention, none of the known nitrogenous wastes, urea, uric acid or creatinine, is responsible. None of these is toxic in relatively large doses; uremia is not produced when animals are fed these substances and their concentrations in the blood raised well above the limits found in renal insufficiency. Bollman and Mann, experimenting upon dogs, implanted the ureters into the intestine. The continued reabsorption of urine caused a rise in the blood urea nitrogen to over 800 mg per cent yet no toxic symptoms resulted. No specific toxic substance (e.g., guanidine or phenolic bodies, as some have suggested) has been definitely demonstrated to be the cause of uremia. The symptoms are now believed to be due rather to a general disturbance in water and electrolyte metabolism resulting in serious abnormalities in the chemical composition of the body fluids. Acidosis and an imbalance of certain electrolytes, especially of sodium, chlorine, calcium, phosphorus and bicarbonate, leading to a disorder in osmotic relationships and in the distribution of intracellular and extracellular water, are looked upon as the principal causative factors.

False or eclamptic uremia is characterized by epileptiform convulsions which are due, apparently, to cerebral edema. This type may occur in the absence of renal insufficiency (as in nephrosis), indeed, it may occur in the absence of renal disease. It is occasionally seen in angioneurotic edema (p. 39) which supports the belief in its being due to cerebral edema. Other signs of increased intracranial pressure, e.g., severe headache, vomiting, slow pulse, hypertension and sometimes choked optic discs, are present. Volhard regards the cerebral edema as due in many instances to intracranial circulatory disturbances—arteriolar constriction with consequent asphyxia and increased permeability of the walls of the cerebral capillaries.

A type of false uremia resembling the foregoing in several respects occurs in arterial hypertension. It is thought to be due to intracranial vascular disturbances alone, i.e., without cerebral edema. Epileptiform convulsions, headache, vertigo, hemianopia, hallucinations, aphasia, temporary paralyses, paroxysmal attacks of dyspnea, or coma with Cheyne-Stokes breathing, may occur. *Hypertensive encephalopathy*, and *cerebral vascular crisis* are terms also applied to this condition.

VII ACIDOSIS

Acidosis is frequently present in the later stages of chronic glomerulonephritis. This is not sur-

prising, since the insufficiency of the kidney involves all its functions, and one of the most important of these is the conservation of base and the regulation of the acid-base balance of the plasma. The fall in the ammonia/acid ratio found in nephritic conditions (p 461) indicates a reduction of ammonia production from glutamine. The glutaminase content of the kidney is reduced (p 458). Acids must then be excreted to a greater extent in combination with fixed base which is therefore lost to the body. The conversion of the alkaline phosphate to the acid salt occurs in the distal tubules, and it is not improbable that impairment of this function is also responsible for a loss of base. Peters and his associates have found a reduction in bicarbonate reserve in a large proportion of non edematous nephritis cases. The diminished alkali reserve was found to be chiefly due to reduction in total base. Contributory causes were the accumulation of phosphates and sulphates and, when carbohydrate of the diet was deficient, of ketones.

The removal of waste products by extra-renal routes

The removal of urea and other retained products in nephritis by purgation and sweating are time honored measures. Sweating was induced by wrap-

ping the patient in hot, moist sheets or blankets, immersing him in a hot bath, or by steaming. The amount of urea which can be eliminated by purgation is relatively small, and the excretion of waste products induced by sweating is not sufficient to be of much benefit to the patient, no more than 2 grams or so of urea can be removed in 24 hours by this means.

Much more effective methods of reducing the azotemia are now available. These are peritoneal lavage with a dialyzing fluid and the circulation of the blood extracorporeally through an artificial kidney. Intestinal lavage has also been carried out by the introduction of fluid through a tube passed into the second part of the duodenum, and withdrawing the fluid again by suction from another tube inserted into the bowel through an opening made into the appendix. This method has not proved very successful. Peritoneal lavage with a dialyzing fluid composed of 0.6 per cent sodium chloride, together with glucose, bicarbonate and the salts of calcium, potassium and magnesium in the concentrations as they exist in serum, has proved more satisfactory. But the best method is the use of the artificial kidney as developed by Kolff of Kampen in Holland. This artificial kidney has evolved from the original viddiffusion experiments of Abel and his associates on animals, in which an apparatus consisting of a number of celloidin tubes immersed in a dialyzing fluid was utilized. Blood entered the apparatus from an artery and, after passing through the apparatus, was returned to the animal by a vein. This device proved remarkably efficient, despite the fact that the total dialyzing surface furnished by the relatively small number of celloidin tubes was only a small fraction (about 3200 square centimeters) of the glomerular surface of the kidney (see fig 36.2).

The ideal apparatus should have (1) a small capacity, in order to prevent a too great withdrawal of the blood from the patient's vascular system and a significant fall in blood pressure, and (2) a maximum dialyzing surface. Kolff's artificial kidney consists of a continuous narrow cellophane tube (about 40 meters long) wound spirally around a horizontal cylinder, which rotates partly submerged in a bath of dialyzing fluid, at a rate of from 35 to 50 revolutions per minute. The total quantity of blood contained in the cellophane tube is about 500 cc. and the area of the dialyzing surface 24,000 square cm. The urea clearance is from 80 to 100 cc. per minute as

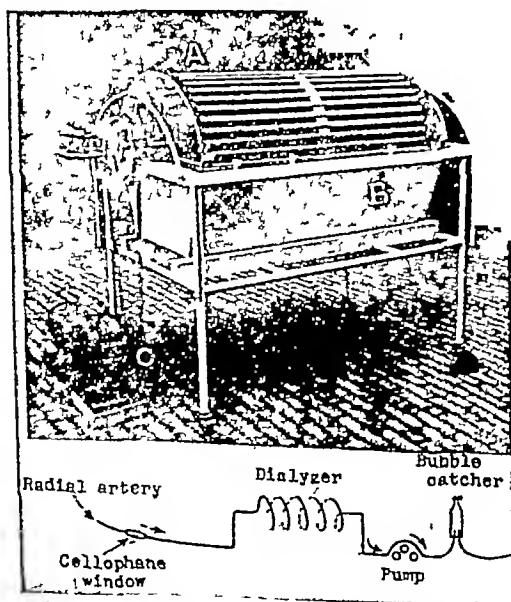


FIG 36.2 A recent model of Kolff's artificial kidney (Courtesy of Dr W J Kolff). Below, diagram of general plan (reversed) of the instrument, A, cylinder upon which the cellophane tubing is wound, B, bath containing dialyzing fluid, C, pump (After Kolff)

compared with 54–75 cc per minute for the normal human kidney. The patient is connected with the system by means of a cannula inserted into the radial artery, and a short length of rubber tubing. The blood enters the cellophane tube under its own pressure and is transmitted to the other end by the rotation of the drum, i.e., by gravity. After passing through a clot and air-bubble catcher, it is returned to a vein by a tube-pump.

This apparatus has proved of especial value in acute uremia, such as may result from anuria due to trauma, as in the "crush syndrome", nephrotoxic chemicals, or following a transfusion of incompatible blood. It helps to maintain the normal composition of the patient's blood until renal function is restored. The dialyzing fluid, about 100 liters in amount, is a solution of 0.6 per cent NaCl, 0.04 per cent KCl, 0.2 per cent NaHCO_3 , and 1.5 to 2.0 per cent glucose in distilled water, it is maintained at body temperature.

Transplantation of a healthy kidney from one person to another, or from a body immediately after death to the living, has been performed on several occasions, but without permanent success. The transplanted kidney survives and functions for a time, apparently normally, but dies and disintegrates after a month or two. Some day it may be found possible to discover and neutralize the antibodies that cause its destruction.

VIII ARTERIAL HYPERTENSION (SEE CH. 16)

MICTURITION

Outline of the anatomy of the urinary tract

The muscular coat of the *ureter* is arranged in three layers, an *external longitudinal*, a *middle circular* and an *internal longitudinal*.

The muscle of the *urinary bladder* is also disposed in three layers, an *external*, a *middle* and an *internal*. The fibers of the external layer run longitudinally, that is, in the long axis of the bladder. The fibers of the middle layer are thinly scattered, they have a circular arrangement and run both transversely and obliquely to the long axis of the viscus. The fibers of the internal layer follow a reticular pattern but for the most part run longitudinally.

The vesical mucous membrane is separated from the internal muscular layer by a submucous coat of loose areolar tissue. The mucosa is thin and, when the bladder is empty or contains only a small quantity of urine, has a corrugated appearance due to the presence of numerous folds or *rugae* which disappear when the bladder becomes distended. The epithelium is of the transitional type. When the bladder is empty the

mucosa shows several cell-layers, but when distended only two strata of cells are to be seen, a deep stratum of cuboidal cells and a superficial one of large squamous cells.

The human bladder has a capacity of from 350 to 450 cc.

The *peritoneum* covers the superior surface of the bladder only.

The ureters pierce the wall of the fundus (or base) of the bladder very obliquely. From a half to three-quarters of an inch of their lower ends is embedded in the vesical muscle. During contraction of the bladder wall these portions of the ureters are compressed by the muscle fibers, thus preventing the reflux of urine as the intravesical pressure rises. The urethral orifice which is situated at the most dependent part of the bladder is guarded by the *vesical sphincter* (*internal sphincter*), formed by the condensation of the muscle fibers of the circular layer. The triangular area marked out by the urethral orifice and the two ureteral orifices is called the *trigone* (*trigonum vesicae*). The muscle of the bladder wall is referred to as the *detrusor urinae*. The male urethra is embraced in its membranous portion by the *sphincter urethrae membranaceus*, a striated muscle which is frequently referred to as the *external sphincter*. The *bulbocavernosus* muscle, which is applied to the urethral bulb and surrounds the corpora cavernosa penis, also exerts a constrictor action upon the urethra. The urethral wall itself contains two layers of smooth muscle, an outer circular and an inner longitudinal, in women (who do not possess an external sphincter of striated muscle) this serves to prevent the escape of urine after paralysis of the internal sphincter. The epithelial lining of the urethra is of the columnar type except near the bladder where it is of the transitional variety, and a short distance from the external urethral orifice where it is stratified and squamous.

FILLING OF THE BLADDER

The ureters exhibit rhythmical peristaltic contractions which travel at a speed of from 20 to 25 mm per second (rabbit) and at a frequency from 1 to 5 per minute, according to the volume of urine formed by the kidney. The peristaltic waves serve to propel the urine from the pelvis of the kidney to the bladder. The urine therefore enters the bladder not in a continuous stream but in separate squirts synchronous with the arrival of the peristaltic waves.

The detrusor muscle exhibits two types of activity, a sustained contraction or tonus, and intermittent contractions.

The bladder in common with other hollow viscera, is capable of adjusting its tone and so of adapting its capacity to changes in the volume of its contents with relatively little alteration in

internal pressure. It thus differs in its behavior from an elastic hollow sphere composed of a non-living material. For example, when a moderate quantity of fluid is run into the bladder through a catheter the intravesical pressure shows a transient rise, due to the tonic resistance offered by the bladder wall, but then declines again to near its previous level as adaptation occurs. As successive volumes of fluid are introduced the curve of intravesical pressure therefore shows a gradual step-like ascent until the bladder contents are unusually large. From then on further additions of fluid cause a much more abrupt rise in pressure. The adaptation is not to a constant pressure (15 cm. of water), as was previously taught, but is always to a pressure a little higher than that existing before the fluid was introduced. In the experiments of Denny-Brown and Robertson upon normal human subjects fluid was run into the bladder 50 cc. at a time. An increase in volume of the contents of the bladder from 50 cc. to 400 cc. caused little change in pressure (fig. 36.3). When completely paralyzed (as during the first two or three weeks following a transverse lesion of the cord) the bladder does not behave in this way, but simply, as an inert elastic bag showing no tendency to respond to an increase in tension upon its walls or to adapt itself to its contents (Holmes), when successive quantities of fluid are introduced the curve of intravesical pressure rises along a smooth line.

Tension is the adequate stimulus for the sensory end organs in the bladder wall. When therefore the bladder becomes distended by the accumulation of urine and the intravesical pressure reaches a certain value, rhythmical contractions of the

detrusor muscle are set up. As the pressure rises further these culminate in the movements constituting the micturition reflex, namely, a strong contraction of the detrusor muscle accompanied by relaxation of the internal sphincter, and followed by opening of the external sphincter. It is usually stated that the reflex occurs at an intravesical pressure of from 15 to 18 cm. of water, but it may be activated by a pressure considerably lower than this. The urine is expelled with considerable force, the pressure within the bladder rising during the contraction of the detrusor to around 130 cm. of water. Since adaptation requires a certain time to take place, the reflex is activated at a lower urine volume than usual if the accumulation of urine is rapid. At the average rate at which urine forms, micturition occurs, unless restrained, after from 250 to 300 cc. have collected.

THE VOLUNTARY CONTROL OF MICTURITION

The act of micturition though essentially reflex in nature is usually initiated by an effort of the will, it also can be voluntarily inhibited or be interrupted at any stage. The desire to urinate is accompanied by a vague feeling in the penis or perineum. The sensation appears when the urine volume is from 200 to 300 cc. If the act is long postponed a feeling of fullness and discomfort culminating in pain results. It is only in the infant or when, as a result of disease, the bladder is isolated from the control of the higher nervous centers, that micturition is a purely reflex act. Under ordinary circumstances in the adult, when the desire to micturate arises, the act is restrained until an opportunity for emptying the bladder presents itself. The restraint is then lifted and the reflex occurs automatically.

The voluntary restraint exerted upon micturition consists, according to Denny-Brown and Robertson, of inhibition of the detrusor with contraction of the external sphincter and perineal muscles. No evidence was obtained by these observers that the internal sphincter was under direct voluntary control, though a reciprocal increase in tone of the sphincter muscle accompanies inhibition of the detrusor. It was found that at a certain urine volume, contractions of the bladder and a consequent rise in intravesical pressure could be readily induced by an effort of the will, contraction of the abdominal muscles was not necessarily associated with the performance of the act. The removal of restraint and the voluntary facili-

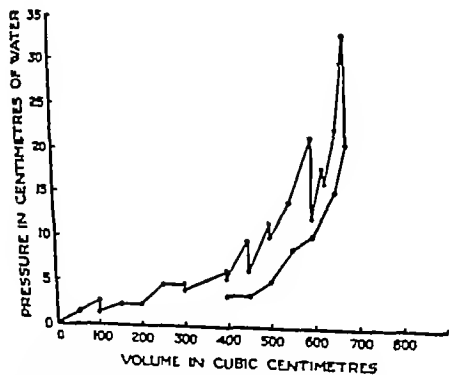


FIG. 36.3 Curve of pressure changes in the human bladder during filling (upper curve) and emptying (lower curve) (After Denny Brown and Robertson)

tation of the spontaneous bladder contractions are considered to be the important factors. At low urine volumes restraint is apparently exercised subconsciously. These observers therefore regard the act of micturition as being normally controlled through variations in voluntary and subconscious restraint of the reflex mechanism. Micturition therefore presents a very unusual feature in that movements innervated by autonomic nerves can be controlled by voluntary impulses. The higher centers from which the inhibitory impulses emanate are situated in the hypothalamus and cerebral cortex, probably in area 4.

Though the abdominal muscles play a non-essential part in micturition, the act, under ordinary circumstances, is started by the contraction of these muscles, and it is well known that the bladder can be emptied though it contains only a few cubic centimeters of urine. The flow of urine is also accelerated during micturition by the rise in intra-abdominal pressure induced by the voluntary contraction of the abdominal muscles. Relaxation of the muscles of the perineum occurs as an associated movement at the commencement of micturition. After the bladder has been emptied the bulbocavernosus muscle (ejaculator urinae) contracts and expels urine which had been left in the urethra.

Contractions set up in a distended bladder by sudden increases of intra-abdominal pressure acting upon the viscus, as in coughing, sneezing, defecation, etc., may, by forcing a little urine past the sphincter into the urethra, cause micturition unless a strong effort of the will is exercised. Psychic influences may also induce bladder contractions which evoke the act unless opposed by restraint. The lifting of voluntary inhibition, unless the bladder is fully distended, may on the other hand be prevented and the power to micturate be temporarily lost, as when a shy or nervous examinee is asked for a specimen of urine in the presence of another person.

THE REFLEX MECHANISMS OF MICTURITION

Barrington describes six integrated reflexes as constituting the act of micturition in the cat, namely

(1) *Contraction of the detrusor* evoked by distending the bladder, the afferent and efferent limbs of this reflex are in the pelvic nerves, its center in the hind-brain. Contraction of the detrusor is accompanied by reciprocal relaxation of the internal sphincter.

(2) *Contraction of the detrusor* caused by running fluid

through the urethra. The afferent pathway for this response is in the pudendal (pudic) nerves, its efferent limb in the pelvic nerves and its center in the hind-brain. Through this reflex the contraction of the detrusor caused by the first reflex is sustained until the bladder is completely emptied.

(3) *Contraction of the detrusor* (transient and weak) when the proximal portion of the urethra is distended. The hypogastric nerves contain both afferent and efferent paths for this reflex, its center is in the sacral part of the cord.

(4) *Relaxation of the external sphincter* when fluid passes along the urethra. Afferent and efferent fibers are in the pudendal (pudic) nerves, the center is in the sacral part of the cord.

(5) *Relaxation of the external sphincter* when the bladder is distended. The afferent path is in the pelvic, the efferent in the pudendal nerves, its center in the sacral part of the cord.

(6) *Relaxation of the plain muscle in the proximal third of the urethra* caused by distending the bladder. Both afferent and efferent paths of the reflex are in the pelvic nerves, its center is in the sacral part of the cord.

In the normal act of micturition the first of these reflexes, namely, contraction of the detrusor in response to distention of the bladder, brings the others, with the exception of the third, automatically into action. It is questionable whether the third reflex is called into play under ordinary circumstances.

THE INNERVATION OF THE URINARY TRACT

The *ureters* in their upper part receive sympathetic fibers from the renal plexus, in their middle part from the spermatic (or ovarian) plexus and near the bladder from the hypogastric nerves. The sympathetic fibers to the ureters exert a predominantly motor effect, though it appears that inhibitory fibers are also derived from the sympathetic. It is thought by some that the ureter receives fibers from the parasympathetic as well, since certain parasympathomimetic drugs cause motor effects. The existence of a parasympathetic innervation has not been demonstrated anatomically.

The *efferent nerves to the bladder* are the sympathetic and parasympathetic. The *sympathetic* furnishes inhibitory fibers to the detrusor muscle; and motor fibers to the trigone, internal sphincter, and the smooth muscle of the proximal part of the urethra.⁷ These fibers arise from the lumbar spinal

⁷ The sympathetic also causes contraction of the ureteral orifices and of the muscle of the seminal vesicles, ejaculatory ducts and prostate.

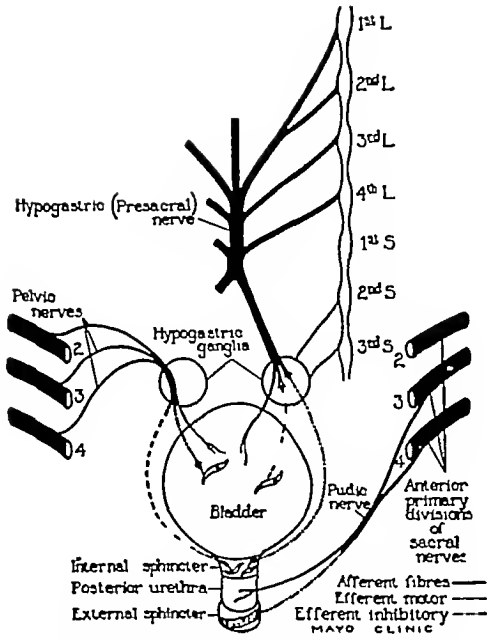


FIG 36 4 To show innervation of the bladder (After Learmonth)

internal sphincter The sympathetic and parasympathetic nerves are therefore reciprocal in their actions The parasympathetic fibers arise from the second and third sacral spinal segments and to some extent from the first and fourth They are conveyed in the pelvic nerves (*nervi erigentes*) to connect with ganglion cells lying in close relation to the bladder wall (see also ch 72) According to Henderson and Roepke, the tonus mechanism is cholinergic (ch 72) but not the phasic contractions

Both sets of autonomic nerves apparently exert a constant influence upon the tone of the detrusor and internal sphincter, the effect of one set being balanced against the other Section of the sympathetic or parasympathetic causes, respectively, an increase or decrease in tone of the sphincter In paresis of the bladder due to injury of the parasympathetic innervation, excision of the pre-sacral nerve is sometimes performed with the object of removing the inhibitory influence of the sympathetic and thus enhancing the action of the pelvic nerves Voluntary control of the detrusor is exerted apparently through the pelvic nerves

The striated muscle constituting the external sphincter is innervated through the pudendal (pudic) nerves

The *afferent* paths from the bladder travel in both *pelvic* and *hypogastric nerves*, those from the urethra in the *pudendal* nerves The afferent fibers essential for the reflex movements of the bladder are contained in the pelvic nerves, those for the movements of the urethra in the pudendals The hypogastric nerves contain no afferents for any of the important reflex mechanisms The sensations set up by distension of the bladder are conveyed in both the pelvic and hypogastric nerves The sensation of pain travels chiefly in the hypogastriacs, but also in the pelvic nerves Excision of the pre-sacral nerve is practised for the relief of vesical pain Tactile and thermal sensations and the sensation of pressure or filling of the bladder are conveyed mainly in the pelvic nerves (Learmonth)

THE NERVE CENTERS GOVERNING MICTURITION

Centers for micturition are situated in the mid-brain, hind-brain and spinal cord The observations cited on page 482 upon the voluntary control of the detrusor muscle also indicate the existence of a center at the cortical level, and electrical excitation of the premotor area causes a rise in vesicle pressure followed by micturition Increase in the

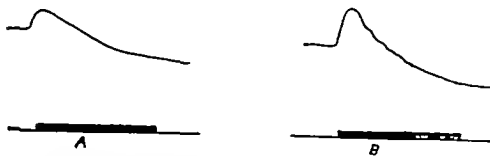


FIG 36 5 Kymographic records showing initial contraction followed by inhibition of the detrusor muscle in response to faradic stimulation of the hypogastric nerves in a dog in which the sacral preganglionic parasympathetic outflow had been interrupted (After Kuntz and Saccomanno)

tone of the bladder wall follows electrical stimulation of the anterior hypothalamic nuclei, stimulation of the posterior nuclei causes inhibition (Beatty and Kerr)

Barrington, experimenting with cats, found that destruction of a small area of the hind-brain (lying ventral to the internal edge of the superior cerebellar peduncle and extending forward from the level of the motor nucleus of the fifth nerve to the anterior end of the hind-brain) temporarily abolished the animal's ability to empty its bladder voluntarily. Destruction of an area in the mid-brain extending from the ventral part of the posterior end of the cerebral aqueduct to just beyond the mesencephalic root of the fifth nerve, was followed by permanent loss of voluntary micturition.

The reflex performance of the act was not, however, impaired. A more extensive lesion in the region of the mesencephalic root caused, in addition, frequency of micturition. Levin and Longworth observed that, in cats, injury to the tegmentum of the mid-brain was followed immediately by hyperactivity of the detrusor muscle to stretch. The capacity of the bladder becomes reduced as a consequence, and rhythmic waves of contraction appear. After a time, this increased activity diminishes and the bladder enlarges again.

The spinal centers lie in the second, third and fourth sacral segments. Only the third, fourth, fifth and sixth reflexes described on page 483 are carried out through the spinal centers.

The descending and ascending paths conveying impulses to and from the spinal centers of micturition are situated in the dorsal half of the lateral column of the cord near its periphery, i.e., in close proximity to the pyramidal tract. The fibers of these paths show extensive crossing in the sacral segments (Barrington).

THE EFFECTS OF NERVE SECTION AND OF CORD INJURIES UPON MICTURITION

Section of the hypogastric (presacral) nerves does not interfere with micturition, these nerves contain neither efferent nor afferent fibers essential for the performance of the act. Incontinence which might be expected to result (since motor fibers are conveyed by this nerve to the internal sphincter) does not occur. Indeed, incontinence does not follow a prostatectomy operation involving destruction of the internal sphincter, for the external sphincter is capable alone of preventing the escape of urine. Frequency of micturition was found by Barrington to follow section of the hypogastrics in the cat, it also occurs as a temporary effect of resection of

the presacral nerve in the human subject. The phenomenon is due, apparently, to the loss of the inhibitory action of the sympathetic upon the tone of the detrusor and the impairment of the ability of the bladder to adapt its capacity to as large a volume of urine as usual.

Section of the pudendal nerves in the human subject is also, according to Learmonth, without any notable effect upon micturition, though of course paralysis of the external sphincter results.

Section of the pelvis causes paralysis of the bladder wall. The detrusor is atonic, the tone of the internal sphincter is raised. Retention of urine with overdistension of the bladder and overflow—dribbling—occurs. Barrington found that the tone of the sphincter diminished after a few days and the animal, apparently experiencing the sensation of fullness of the bladder, assumed the usual position for micturition and performed the act by a contraction of its abdominal muscles.

Severance of the posterior sacral nerve roots is followed immediately by the loss of all the important reflexes of micturition since the afferent paths (through the pelvis and pudendals) are interrupted, the bladder wall is flaccid and the resistance at the internal sphincter is increased. However, after a period of overdistension with overflow incontinence, the bladder may empty automatically at intervals. The sensations, except pain of overdistension (afferent fibers of hypogastrics intact) are retained.

Destruction of the sacral nerves (as in lesions of the cauda equina) or of the spinal centers will interrupt not only motor impulses to the bladder, but also the afferent impulses travelling by the pelvis and pudendals. The bladder is then completely isolated from central nervous control, but, after a period of retention of urine with overflow it may partially expel its contents automatically. The detrusor and internal sphincter act coördinately. Such an action suggests a neural mechanism of some sort. It occurs in animals even after the hypogastrics have been sectioned as well, and time allowed for degeneration to occur, it cannot therefore be due to preganglionic axon reflexes (p. 290). Probably, in such and other instances in which motor impulses from the central nervous system have been interrupted, local reflex arcs through ganglion cells in the vesical plexus or bladder wall are responsible for the automatic action. The latter is only a makeshift for the normal micturition reflex since the bladder is not completely emptied but always retains a part of its contents (residual urine). McLennan draws attention to the fact that the state of the bladder after the afferent limb of the reflex arc is interrupted as in tabes is different from that following destruction of both afferent and efferent limbs. In the former the bladder is lax (hypotonic) and of large capacity, while in the latter it is hypertonic and of normal capacity. The reason for such a difference is obscure.

2. After transection of the cord above the sacral region,

normal micturition cannot occur. Nevertheless, after a variable period of retention with overflow the spinal (sacral) centers assume control and the bladder empties at intervals automatically (see also mass reflex, ch 66). Owing to the absence of the first and second reflexes of Barrington which, as just mentioned, are governed by higher centers, the bladder does not empty itself completely, but always contains a quantity of residual urine.

Bladder sensibility is completely lost following section of the hypogastrics and pelvic nerves or transection of the cord above the entrance of the afferent fibers of the hypogastrics. Injury or disease (e.g., tabes) of the posterior columns involving these fibers together with those from the sacral roots will have a similar effect.

Bilateral lesions of cortical areas 4 and 6 in man abolish voluntary control of the bladder, urinary

incontinence results. Incontinence is seen not uncommonly in the aged, and may occur though no organic lesion is apparent, it is attributed to impairment of voluntary control by the cortical centers.

Nocturnal enuresis, or the passage of urine during sleep, occurs in young children up to an age which varies considerably, and may be attributed to the undeveloped state of the neural mechanism through which the act of micturition is voluntarily inhibited. But enuresis during sleep sometimes persists beyond the usual time and even to adult age. It may then be associated with some definite organic abnormality, e.g., in the lumbosacral vertebrae, e.g., spina bifida occulta, but nocturnal enuresis in the adult is usually related to some functional neurological disorder, or to psychological factors. Evidence of pelvic autonomic dysfunction or psychiatric conditions in the subject's immediate family are present in a large proportion of such persons. In others there is definite mental deficiency.

SECTION V. DIGESTION

By N B T

CHAPTER 37

THE SALIVARY GLANDS AND THE SECRETION OF SALIVA

General description and structure of the glands

Saliva is secreted mainly by three paired masses of cells—the *submaxillary*, *sublingual* and *parotid glands*. There are also small glands scattered over the buccal mucous membrane, which secrete a mucoid fluid. The cells of the glandular tissue are aggregated into a great number of small groups, with the individual cells arranged in a single layer around a small central cavity or alveolus. The cells are more or less wedge-shaped with their apices converging toward the central cavity and their bases directed outwards. They discharge their secretion into the alveolus which is drained by a fine duct. Ducts from neighboring alveoli join to form ducts of larger caliber which unite again to form still larger trunks, until finally through a succession of junctions and the formation of channels of ever increasing size, the secretion flows into the mouth by a single large duct, in the case of the submaxillary and parotid glands, or by several of medium size, in the case of the sublingual. The general arrangement of the ducts reminds one of the stem branchings of a bunch of grapes—the rounded alveoli at the ends of the finest channels corresponding to the grapes. Glands showing such a pattern are, therefore, termed *racemose*. The ducts are named, in accordance with their location, *intralobular* or *interlobular*. The former drain a single alveolus or a group of alveoli. The latter lie between the lobules each being formed by the union of a number of intralobular ducts. The narrowed part of the intralobular duct lying nearer to the alveolus is known as the *isthmus*, or more usually the *intercalary duct*.

There are two main types of salivary cells—*serous* and *mucous*. The alveoli of the *parotid gland* are composed entirely of one type—the serous. This type of cell secretes a thin watery fluid. The *submaxillary gland*, on the other hand, in man, is of the mixed type, some of its alveoli are composed entirely, like the parotid, of serous cells, while others contain only cells of the mucous type. The two varieties of alveoli are indiscriminately intermingled with one another, either singly or in clusters. The mucous cells secrete a thick, viscid juice rich in mucin. In the *sublingual*, the alveoli are predominantly of the mucous type though a few serous alveoli may also be seen. In both the submaxillary and sublingual glands, particularly the former, the two types of cells may also be found in a single alveolus.

The serous cells then appear as crescent-shaped elements—*demi-lunes* or crescents of Gianuzzi—pressed against the periphery of the alveolus between the limiting membrane and the larger, more numerous mucous elements (fig 37 1). Cells in the walls of the intercalary and intralobular ducts, known as *isthmus* and *rodged epithelial cells*, respectively, are believed also to have a secretory function, but the exact nature of this is unknown.

The *parotid secretion*, as naturally follows from the character of its cells, is thin and watery in nature, it has a low content of organic material. It is delivered into the mouth through the *duct of Stensen*, which opens upon the inner surface of the cheek opposite the second molar tooth. The duct of the submaxillary—*duct of*

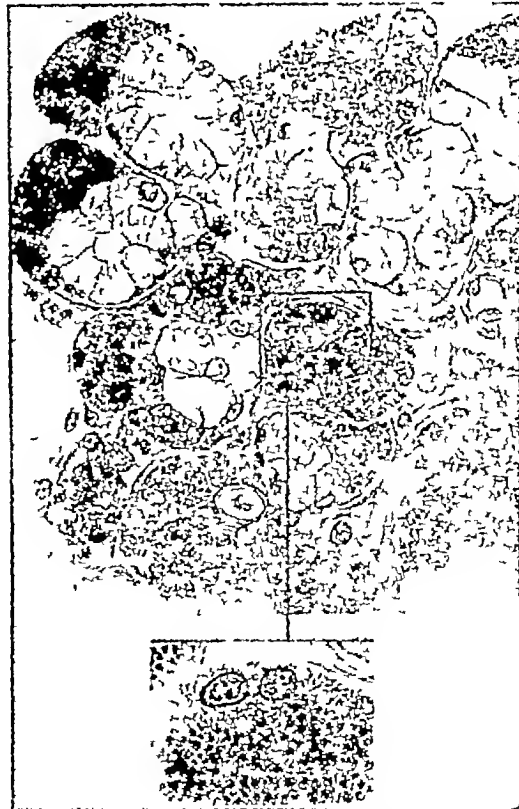


FIG 37 1 Human submaxillary gland showing both mucous and serous groups of cells. Inserts are views with oil immersion, zymogen granules are clearly shown (After Stormont, from Cowdry's *Special Cytology*)

Wharton—runs from the substance of the gland which lies under shelter of the mandible, and opens upon the floor of the mouth to one side of the frenum of the tongue. The secretory ducts of the sublingual—the *ducts of Rivinus*—are several fine tubes which open beside the frenum. The secretion from the submaxillary gland may be either thin and watery (serous) or thick and viscid (mucous), the character at any time depending, as shall be shown later, upon the nature of the secretory stimulus. The *sublingual secretion*, on account of its scarcity in serous cells is usually of the mucous type.

MINUTE APPEARANCE OF THE SECRETORY CELLS The cytoplasm of either the serous or mucous cell is not homogeneous but shows a granular structure, due to the presence of minute droplets of a colloidal material. The granular appearance is different in the two cell types. In the case of the serous cells they are fine, and when the gland is in the resting state, load the cytoplasm to such an extent that the nucleus is almost obscured. These droplets are believed to furnish the enzyme of the secretion, and are consequently termed *zymogen granules*. In the case of the mucous cells coarser *mucinogen granules* are seen. From these are derived the mucin which gives to the secretion of these cells its slimy character. When the glands enter upon secretory activity, the granules—whether zymogen or mucinogen—becomes less numerous as a result of their extrusion into the alveolus. After a prolonged period of secretion only a few remain in the region of the cell bordering the cavity of the alveolus, the rest having been discharged along with the water and other constituents of the juice. After a period of rest they accumulate again and gradually fill the cell.

THE CONTROL OF SALIVARY SECRETION

Two methods are at the disposal of the body for the control of secretion. Glandular cells, in general, may be excited either through the mediation of nerves or chemically, by the action of hormones (ch 57). Though there are exceptions, a quick response calls for the nervous type of control. On the other hand, when rapidity of response is not a necessity, hormone control is employed. It will be seen later that some glands are controlled by both methods. Usually in such instances nervous impulses initiate the secretion and hormones then maintain it over longer periods. It would appear that nervous energy is economized whenever it is possible for the secretion to be carried out efficiently by hormonal means. In the case of the salivary secretion a rapid response, in man and the carnivora at any rate, is obviously essential, since the food remains for such a short time in the mouth. Nervous mechanisms, therefore, are relied upon to bring about salivary secretion,

there is no evidence of a specific hormonal mechanism, as shown years ago by Langley, and subsequently confirmed by others, the salivary glands fail to respond to acid in the mouth after section and degeneration of the secretory nerves. The salivary cells, however, are by no means unsusceptible to chemical influences, for a large number of substances, e.g., drugs and abnormal metabolic products, reaching them through the blood stream are capable of influencing their activity. The two types of cell respond selectively to different drugs (p) 493). Even though the salivary glands are under nervous control the immediate excitant of the gland cells is believed to be a chemical substance (humor) liberated at the nerve endings (see ch 72).

Innervation of the salivary glands

The glands are supplied with secretory nerves from two sources—the *bulbar* and the *thoracolumbar* divisions of the *autonomic nervous system* (ch 72). The bulbar centers consist of a group of nerve cells which runs forward from the anterior end of the glossopharyngeal nucleus to the sensory nucleus of the facial nerve. The anterior (rostral) part is termed the *superior salivatory nucleus*, and governs secretion by the submaxillary gland, the posterior (caudal) part, called the *inferior salivatory nucleus*, controls secretion by the parotid gland.

The submaxillary and sublingual glands receive secretory impulses through the *chorda tympani nerve*. These fibers have the following course. Arising from the *superior salivatory nucleus* they leave the brain in the *nervus intermedius* of *Wrisberg* (ch 66). They pass without interruption through the genicular ganglion of the facial nerve and descend with the facial to the point where its *chorda tympani* branch is given off. They enter this nerve which upon approaching the cavity of the mouth joins the lingual nerve. In the floor of the mouth the secretory fibers leave the lingual again to make connections with the nerve cells of small ganglia. From these, postganglionic fibers arise which terminate by fine arborizations around the secretory (serous) cells (fig 37.2). The ganglia furnishing fibers to the submaxillary gland are small and numerous, and are situated in the hilum of the gland. The sublingual gland contains no ganglion cells, it receives its postganglionic fibers from a small ganglion—the *submaxillary ganglion*—situated in the course of the chorda fibers just beyond their point of separation from the lingual and before they enter the gland (see diagram fig 37.3).

The *sympathetic supply* (postganglionic fibers) is de

rived from the superior cervical ganglion. The fibers reach the gland via the plexuses of the external carotid artery and its branches. The preganglionic fibers arise from the upper one or two thoracic segments of the cord. Fine filaments end in the secretory (mucous) cells.

The chorda tympani also carries *dilator fibers* to the blood vessels of the gland, *vasoconstrictor fibers* are derived from the sympathetic.

The bulbar fibers to the *parotid* pursue the first part of their course in the glossopharyngeal, but they follow a devious path before finally terminating around the gland cells. They arise in the medulla from the *inferior salivatory nucleus*. At the jugular foramen they separate from the glossopharyngeal (petrous ganglion) in its *tympanic branch* (*Jacobson's N*), and after passing into the trunk of the *small superficial petrosal nerve* are conveyed to the *otic ganglion*. There they communicate with ganglion cells from which postganglionic fibers arise. The latter are transmitted by the *auriculo-temporal* branch of the fifth nerve to the gland cells. The parotid also receives sympathetic secretory fibers. The cell stations of the latter as in the case of the submaxillary and sublingual lie in the superior cervical ganglion. The dilator fibers to the blood vessels travel with the bulbar fibers. The vasoconstrictor fibers are furnished by the sympathetic, and follow the blood vessels into the gland.

EXPERIMENTAL STIMULATION OF THE SECRETORY NERVES

The nature of the secretory response which follows the electrical stimulation of one of the efferent salivary nerves depends upon which of the two types of nerve is chosen for the experiment. If the chorda tympani is stimulated, a profuse watery juice, poor in solids, results, the vessels of the gland dilate and the blood flow through it increases. The oxygen consumption of the gland is increased two- or three-fold. If a manometer is placed in the salivary duct and another in the carotid artery it will be found that when the chorda tympani is stimulated continuously the secretory pressure ultimately rises above the blood pressure. This fact and the increased oxygen consumption of the gland during activity show that the production of saliva is a true secretory process, and not simply due to filtration of fluid from the blood stream.

Stimulation of the auriculo-temporal nerve (parasympathetic fibers) to the parotid also causes an abundant watery secretion. Stimulation of the sympathetic fibers to the submaxillary or sublingual gland causes the secretion of a smaller quantity of a thick mucinous saliva. No apparent

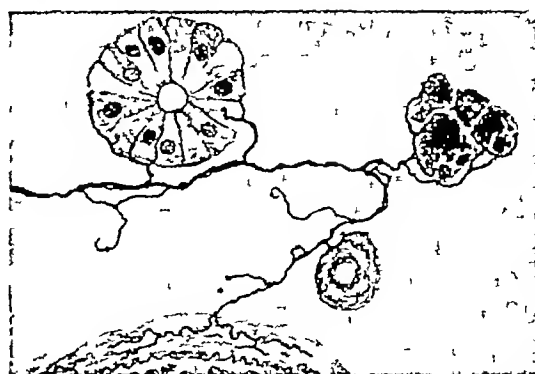


FIG 37.2 Showing distribution of parasympathetic secretory nerves to cells and blood vessels of submaxillary gland of rabbit. The serous cells are believed to be supplied by the parasympathetic, the mucous cells by the sympathetic. (From Stormont in Cowdry's *Special Cytology*)

secretion follows stimulation of the sympathetic supply to the parotid, though upon microscopical examination of the gland after such stimulation a reduction in the zymogen granules of the cells can be observed, which shows that sympathetic stimulation is not without a secretory effect. Stimulation of the cerebral cortex in the dog behind and below the cruciate sulcus causes the secretion of saliva.

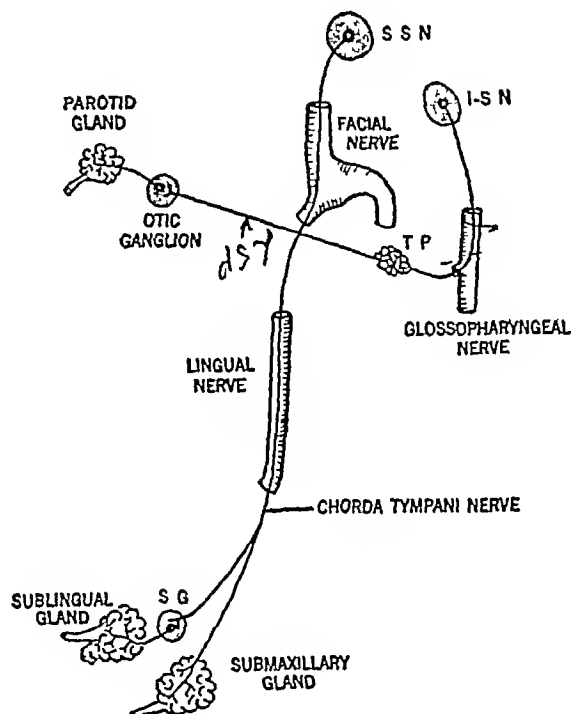


FIG 37.3 Diagram of the parasympathetic nerve supply of the salivary glands. S-S-N, superior salivatory nucleus, I-S-N, inferior salivatory nucleus, S-G, submaxillary ganglion, T-P, tympanic plexus.

THE REFLEX SECRETION OF SALIVA (ARTIFICIAL STIMULATION)

The salivary glands can be readily excited by the artificial stimulation of afferent nerves supplying structures in the mouth. The reflexes are brought about through the salivary centers. The *efferent* limbs of the reflex arcs—the secretory fibers of the chorda tympani and the tympanic branch of the glossopharyngeal respectively—have been considered above. The *afferent* pathways are along fibers contained in the trunks of the chorda tympani and glossopharyngeal nerves, in the lingual, buccinator and palatine branches of the fifth nerve and in the pharyngeal branch of the vagus. The fibers of the chorda tympani subserving the sensation of taste are distributed to the anterior two-thirds of the tongue. They arise from cells in the genicular ganglion, the central processes of the ganglion cells enter the pons in the *nervus intermedius* of *Wrisberg* to make connections with cells in the tractus solitarius from which impulses are relayed to the superior salivatory nucleus. The afferent fibers of the glossopharyngeal nerve concerned in the salivary reflex, carry sensations of taste from the posterior third of the tongue. They arise from cells in the petrous ganglion. The central processes of these cells enter the medulla to make connections through the tractus solitarius with the inferior salivatory nucleus (see also p. 488). The lingual fibers (cells of origin in Gasserian ganglion) furnish the general buccal mucosa with common sensation—touch, pain, etc. Secretion of saliva may be induced by stimulation of the central end of any one of these three groups of afferent fibers. It can also be brought about by stimulating sensory nerves in other situations. For example, experimental stimulation of the central end of the vagus, sciatic or indeed of practically any sensory nerve of the body may cause salivation. According to some, stimulation of afferent nerves of the gastric mucosa is particularly likely to initiate a reflex secretion into the mouth. In disease, stimuli arising in the esophagus may cause profuse salivation (esophago-salivary reflex). Stimulation of trigeminal terminals in the nasal mucosa will also evoke a secretion of saliva (*Babkin*).

By means of the string galvanometer, an electrical change of about 2.5 millivolts can be demonstrated in the salivary gland during secretion. One electrode is placed on the surface of the gland and the other on an indifferent part of the body. The electrical changes during secretion have been studied in the submaxillary

gland by *Langenskiöld*, who finds that the electrogram obtained during chorda stimulation differs from that caused by stimulation of the sympathetic nerve supply. The former shows a sharp positive deflection, the latter a prolonged wave below the base line.

THE SECRETION OF SALIVA UNDER NATURAL CONDITIONS

In the normal life of the animal the secretion of saliva is brought about reflexly in two ways, either through (1) the stimulation of the nerves of the mouth by the presence therein of food or other substances, or (2) by the stimulation of some organ of special sense other than that of taste. The former type of reflex is termed *unconditioned* or *inherent*, the latter, *conditioned* or *acquired*. A reflex of one type does not, of course, exclude the other, and as a matter of fact both are called into play together under ordinary circumstances.

(1) *The unconditioned salivary reflex*

Materials placed in the mouth call forth, after a short latent period (two or three seconds), a secretion of saliva which varies in *quantity* and *quality* with the physical and chemical nature of the substance introduced. The effects which sensations of taste produce upon the secretion of saliva are well known. Among edible substances, those, generally speaking, which are the most palatable or arouse the sensation of taste with the greatest intensity, are the most potent salivary stimulants. Materials that are entirely inedible will, if unpleasant to the taste—acids especially—cause profuse salivation. In these instances the secretion depends mainly upon the stimulation of the taste fibers, and the stimulus is chemical in nature. But we have seen that stimulation, not only of the taste fibers (chorda tympani and glossopharyngeal) but of the fibers endowing the mucosa of the mouth with common sensibility (lingual nerve) as well, will produce a salivary flow. So, materials such as dry sand, inedible powders, whether soluble or insoluble, or any other material which is capable of stimulating these endings in a purely physical way will evoke a secretion. The mere movements of the jaws and of the tongue over the mucosa of the mouth will have such an effect, though there is no material present in the mouth. A secretion occurs when any substance is chewed, whether or not it is edible or possesses taste. The chewing of india rubber (or gum) for instance, the manipulations of the dentist, the contact of his instruments with the oral mucosa or the grinding of a tooth are familiar and potent causes of salivation.

The remarkable *adaptability* or *purposeful character* of the salivary reflex has been remarked upon by Pavlov. The physical and chemical qualities of the juice, as well as its quantity, are adapted to the physical or chemical characters possessed by the particular substance initiating the reflex. For instance, if clean pebbles be placed in a dog's mouth, they are expelled—merely allowed to drop out of their own weight. No secretion or very little occurs, since none is required, but if the stones are crushed and given as a powder, a profuse watery salivation follows to rid and cleanse the mouth of the useless material. The juice in this instance is poor in organic material and resembles that obtained upon electrical stimulation of the chorda tympani nerve or of the parasympathetic fibers to the parotid. Strong acid produces an abundant saliva which, according to Pavlov, is relatively rich in protein, this exerts a buffer action which reduces or annuls the injurious effect of the acid. The salivary response to the various foods is also adapted to their peculiar qualities. A chunk of meat, if given to a dog, is very quickly swallowed. Under the circumstances what is most required of saliva is a lubricant action. Accordingly, a highly viscous juice, rich in mucin, is produced. If the meat be first dried and powdered, or if dry biscuit be fed to the animal the secretion is characteristic of parotid or chorda saliva—e.g., watery and abundant, but poor in mucin. Milk evokes the secretion of a saliva rich in mucin, and foods, in general, produce a saliva rich in organic material—mucin and ferments—while inedible substances tend to call forth a more watery juice. These adaptations are much less pronounced in man.

The conditioned or acquired reflex

The secretion which flows into the empty mouth when "the mouth waters" is the result of a conditioned reflex. The stimulus which initiates such a reflex is not applied to the nerves of the mouth but is received by one or other of the organs of special sense, particularly those of sight and smell. A conditioned reflex may also be elicited through the sense of hearing or through sensory impressions arising from stimuli applied to the skin. In brief, a conditioned reflex is one in which the cerebral centers play an essential part, and in which training and experience are the basis for the development of the reflex process. Conditioned reflexes are taken up in detail in chapter 69.

THE REACTION, QUANTITY AND COMPOSITION OF SALIVA

Human mixed saliva, according to the investigations of Starr, is slightly acid in reaction. In 86 per cent of a large series of normal persons the pH was found to vary between 6.35 and 6.85. The lowest pH found was 5.75 and the highest 7.05. Salivary reaction is dependent

mainly upon the relative concentrations of free and combined CO_2 , that is, upon the ratio $\frac{[\text{H}_2\text{CO}_3]}{[\text{NaHCO}_3]}$. So, in order that the true pH value be obtained the juice must be collected without loss of CO_2 . The older figures in the literature are too high (7.50 to 8.00) since this precaution was not taken. The hydrogen ion concentration of the saliva was found to vary directly with the CO_2 content of the blood. This means that when the CO_2 tension in the blood is high, more CO_2 finds its way into the salivary secretion to lower its pH, and vice versa. Forced breathing causes a lessened amount of CO_2 in the saliva, and consequently a rise in its pH. On the other hand, conditions associated with a retention of CO_2 in the blood increase the loss of the gas in the saliva. Ingestion of NaHCO_3 , while it reduces the acidity of the urine, increases that of the saliva, since a rise in the CO_2 tension in the blood results.

In man the amount of saliva secreted in 24 hours amounts to from 1000 cc. to 1500 cc., but the output of the resting gland is only about 0.25 cc. per minute. The cow secretes some 60 liters daily. Ordinary mixed saliva contains about 99.5 per cent of water and 0.5 per cent total solids. It has a specific gravity between 1.002 and 1.012. Its main constituents are as follows:

I *Salts* (approximately 0.2 per cent)

Sodium and potassium chloride
Sodium bicarbonate
Acid and alkaline sodium phosphates
Calcium carbonate and calcium phosphate
Potassium sulphocyanate

II *Gases*

Carbon dioxide, oxygen and nitrogen

III *Organic substances*

Ptyalin (salivary amylase), maltase, and lysozyme

Serum albumin and globulin

Urea, uric acid, creatine and amino acids

Mucin, mainly in the submaxillary and sublingual secretions

Glucose is absent, normally, and even in diabetes, none or only small amounts are found in the saliva.

The *bicarbonates* and to some extent the *phosphates* act as "buffers." The buffering action of saliva is such that its pH remains constant under all ordinary conditions and even though relatively strong solutions of acids or alkalis are introduced into the mouth, the reaction, except for a short period, is not altered. The "acid mouth", so glibly spoken of in dentifrice advertisements as an abnormal condition, and the possibility of altering the reaction of the oral secretions are figments of the imagination. Attempts to change the pH of the saliva experimentally by flooding the mouth with acid or alkaline solutions are followed by contrary effects upon the salivary reaction. Acid solutions cause a rise in pH, alkaline solutions a fall. The changes last for about 10 minutes, the reaction then returns to normal.

The *calculus* are necessary for the activation of the amylase. The *calculus* salts which are soluble in acid but insoluble in alkaline media tend to be thrown out of solution when the pH rises. The carbonate and phosphate of calcium may be deposited in the form of concretions (*calculus calculus*) within the crevices or, in combination with organic material, may be laid down upon the teeth as "tartar" at a high salivary pH and a juice rich in mucus are believed to be conducive to tartar deposition and the development of *calculus*. The *hyperuricæmia* (KSCN) is an extremely potent and is probably formed within the body from CN radicals derived from the metabolism of protein. Its production and excretion are thought to represent a detoxicating mechanism. It is said to be in excess in the saliva of habitual smokers, and has been shown by Sullivan and Dawson to be noticeably reduced during the course of *pellagra*, but returns to normal value during convalescence from this disease.

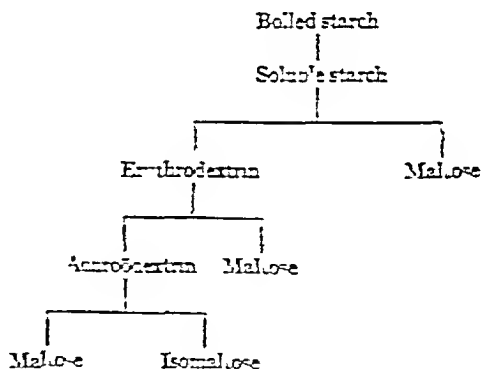
The bactericidal enzyme known as *lysozyme* has a caustic action upon staphylococci, streptococci, meningococci and other microorganisms. It is also present in the lacrimal secretion and is widely distributed in animal tissues and in egg white in which it is closely associated with the protein *albumin* (p. 756). It has a mucolytic action resembling that of hyaluronidase (p. 312 and 391) and probably owes its antibacterial activity to its ability to dissolve the polysaccharide in the capsule of the microorganisms.

FUNCTIONS OF SALIVA

(1) **DIGESTIVE.** The starch molecule is acted upon by ptyalin and split into smaller molecules of the disaccharide maltose. The rapid passage of the food through the mouth precludes the possibility that it is acted upon here by the saliva to any important extent. Whether starchy food after its thorough impregnation with saliva undergoes any significant degree of digestion in the stomach has been debated. The salivary amylase requires for its activity an alkaline, neutral or but faintly acid medium. It was, therefore, thought that the highly acid gastric juice would prevent or soon terminate salivary digestion. It has been shown, however, that the latter part of the meal, which usually contains the carbonate, may remain in the fundus of the stomach protected for some time from the acidifying action of the gastric juice by a layer of food ingested previously. For this reason it is likely that under favorable circumstances considerable digestion of starch is accomplished during this period. Bergam found that 76 per cent of the starch of mashed potatoes was transformed into maltose in the human stomach.

When boiled starch is placed in a test tube with

mixed human saliva and kept at body temperature, a slow conversion of the starch into maltose takes place. The chemical change occurs in a series of stages which may be distinguished by the manner in which the product of each reacts with iodine. Iodine gives a characteristic blue color with boiled starch. A short time after the saliva has commenced to act, a physical change may be seen to have occurred in the starch. It loses its opalescent appearance and becomes soluble, though it still gives the blue color with iodine. After a short time the starch becomes partially broken down and converted into a dextrin which now gives a red color with iodine, and on this account is known as *erythro-dextrin*. Small amounts of maltose may also be detected. Still later, no color reaction occurs upon the addition of iodine, a colorless dextrin—*achro-dextrin*—has been formed. Finally the starch is entirely converted into maltose and isomaltose. In the final stage traces of glucose may also appear due to the presence in the saliva of maltase in low concentration. The following scheme illustrates these changes



Ptyalin has no action upon cellulose and for this reason the starch must be cooked in order that the cellulose envelope surrounding the starch grains may be broken. Boiling also causes hydration of the starch molecule itself, and renders it more easily attacked by the amylase.

(2) **PREPARATION OF THE FOOD FOR SWALLOWING BY ALTERING ITS CONSISTENCY.** This is one of the most important functions of the saliva—the food is moistened, thus enabling it to be rolled into a plastic mass, and given a lubricant coating. Claude Bernard showed that a horse with a parotid fistula had the utmost difficulty in swallowing dry hay or oats.

(3) **SOLVENT ACTION.** Taste is a chemical sense.

All solid substances, in consequence, in order that they may stimulate the taste buds must be dissolved in the saliva

(4) **CLEANSING ACTION** The constant flow of saliva exerts a very necessary cleansing effect. The mouth and teeth are rinsed and kept comparatively free from food residues, shed epithelial cells, foreign particles, etc., in this way the saliva inhibits the growth of bacteria by removing material which may serve as culture media, it also, by virtue of its lysozyme content, is directly bacteriocidal. One has but to consider the foul condition of the mouth in certain fevers when the salivary secretion is suppressed, in order to realize how important are its cleansing and bacteriocidal properties. Then, decomposing organic material swarming with bacteria (*sordes*) collects upon the teeth and lips, and must be removed by artificial means

(5) **MOISTENING AND LUBRICATING ACTION** The saliva by moistening and lubricating the soft parts of the mouth and lips keeps them pliable and resilient for the purposes of articulation. Frequent sips of water are almost essential for some public speakers, in whom as a result of evaporation from the mouth during speech, the supply of saliva is insufficient.

(6) **EXCRETORY** Many substances both organic and inorganic are excreted in the saliva. Drugs such as mercury, potassium iodide, lead, etc., when introduced into the body are excreted in part by the saliva. Severe inflammation of the oral mucosa (stomatitis) may be caused by the excretion of excessive amounts of mercury by this route. The blue line on the gum margins in lead poisoning is due to the metal having been excreted in the saliva, and deposited as the sulphide. The sulphur is provided by organic material contained in the tartar formed on the bases of the teeth. For this reason the discoloration of the gum does not occur where teeth are absent. In chronic nephritis the saliva contains a high percentage of urea, and sugar sometimes appears in severe diabetes, in parathyroid overdosage the calcium concentration of the saliva is elevated. Several types of micro-organisms, some intensely virulent, e.g., the virus of hydrophobia and anterior poliomyelitis are excreted in the saliva. The latter disease has been reproduced in monkeys by injecting the saliva of an infected person.

In this connection it may be added that mumps, which is usually looked upon as a specific inflammation

of the parotid gland, is more likely a general disease, since other organs, e.g. ovary, testicle, cerebral meninges, etc. unconnected in any way with the salivary glands are often seriously affected. The parotid inflammation is probably incidental and results from the passage of the infectious agent through the gland into the saliva.

(7) **THE RÔLE PLAYED BY THE SALIVARY GLANDS IN THE REGULATION OF THE WATER BALANCE OF THE BODY** When the water content of the body is adequate the saliva is secreted continuously into the cavity of the mouth either by the main glands or by the innumerable small mucous glands that are scattered over the surface of the buccal mucosa. When, however, large quantities of fluid are lost from the body either through the sweat, bowels, kidneys, evaporation from the lungs, through loss of blood or when the water intake is curtailed, the salivary glands, in common with the other tissues, are subjected to the dehydrating effect, salivary secretion is suppressed. Drying of the oral mucous membranes, and the consequent stimulation of afferent nerves of the mouth and pharynx arouse the sensation of thirst (p. 604). Thirst may be looked upon as an essential part of a protective mechanism against the depletion of body fluid. It serves to warn the individual that the body's water supplies require to be replenished.

The actions of drugs and chemicals upon salivary secretion

Serous secretion is stimulated by adrenaline, and ephedrine, acetylcholine, muscarine, pilocarpine, physostigmine (eserine), and histamine increase the mucus secretion. Atropine which antagonizes acetylcholine, and ergotamine which paralyzes sympathetic effects, inhibit secretion. Quinine paralyzes the effects of both nerves.

Disturbances of salivary secretion

Permanent suppression of the salivary secretion—*xerostomia* or *aptyalism* as it is termed—is an unusual condition, little is known regarding the mode of its production. Temporary suppression of salivary secretion is more common and occurs in emotional states and in fevers, or as already mentioned, when the water content of the tissues is lowered. Excessive salivation or *ptyalism* is not unusual and is often particularly troublesome in pregnancy, its cause in the later state is unknown, it is possibly of reflex origin, or due to some metabolic product acting in a drug like manner upon the gland cells or the secretory nerves.

As a result of irritation of the gastric mucosa, in duodenal ulcer, or in lesions of the esophagus, such as

carcinoma or spasm of the cardiac sphincter, salivation occurs as a reflex phenomenon, and may be pronounced (esophago-salivary reflex). The latter reflex is usually readily elicited in a normal person by the passage of a stomach tube or an esophageal sound. Since the glands respond to mechanical stimuli, painful or otherwise, the salivation associated with abnormal conditions in the mouth, e.g., a carious tooth, carcinoma of the tongue, etc. is not surprising.

When the reflex secretion is the result of stimuli arising in the stomach, esophagus or duodenum, and is excessive, the saliva may, without the individual's knowledge, pass down the esophagus and collect above the cardiac sphincter. The secretion occurs as a rule shortly after a meal and a short time later a large quantity of fluid may have accumulated, it may then be brought into the mouth in one or two gushes without any vomiting effort or even nausea. The condition is spoken of as *water-brash*.

The form and structure of the teeth

A tooth consists of a *crown*, a *neck* and a *root*. The crown is the part projecting beyond the gum, the root fits into the socket or *alveolus* of the jaw bone. The neck is the junction between crown and root, and, normally, lies just below the gum. Three kinds of tissue, *enamel*, *dentin* and *cementum*, compose the hard portion of the tooth. The cavity occupying the center is filled with a soft substance, called the *pulp*, composed of connective tissue fibers and a gelatinous matrix in which the nerves and blood vessels are embedded (fig 374).

The enamel, which is of epithelial origin, is a white, translucent, and very hard material—the hardest in the body. It is composed of calcium salts (95 per cent) in the form of apatite crystals (ch. 60), with from 3 to 5 per cent of organic material. Its structure consists of thin prisms or rods running through its entire depth, perpendicular to and resting on the dentin, the prisms are cemented together by protein material or one containing protein. The dentin surrounds the pulp

cavity except at the apex of the root where the nerves and vessels enter, it resembles bone chemically and structurally, but is harder, and contains numerous canaliculi—the *dentinal tubules*—which radiate from the pulp cavity. In the pulp lying against the dentine are large elongated cells with a radial and epithelium like arrangement, they are called *odontoblasts* and send fine processes (fibers of Tomes) into the overlying dentinal tubules. The cementum surrounds the root outside the dentin, it is a bony material, but (except in old age), lacks Haversian canals. It is penetrated by bundles of coarse fibers (Sharpey's fibers) derived from the *periodontal membrane*, this membrane lines the alveolus, for which it serves as periosteum, and acts as a close bond between the cementum and the bone. Bone cells are embedded in the cementum near the apex of the root. The viability of the cementum is dependent upon the integrity of the periodontal membrane, it undergoes necrosis when the latter is destroyed.

Like bone, tooth structure is in a continuous state of flux, minerals, calcium and phosphorus being continually removed and replaced.

Dental decay, or dental caries

Calcium salts, which constitute such a large proportion of dental tissue, are soluble in acid, an increase in the acidity of the saliva was thought, therefore, to favor decalcification of the teeth, and in this way prepare them for the inroads of microorganisms. But the saliva is normally slightly acid, and any significant increase in its acidity has not been demonstrated in caries (Geis).

Bacterial decomposition of carbohydrate food with the production of lactic acid, especially in regions where food is likely to collect, as in crevices between the teeth and at the margin between the tooth and gum, has long been recognized as an important element in tooth decay. Sugar, especially if highly refined, is undoubtedly potent in increasing susceptibility to the disease. Sugar has been shown to facilitate the penetration of the H ion into the enamel.¹

Another local factor which has been suggested as predisposing to caries is the mucin content of the saliva. A juice rich in mucin flows less freely and as a result of its high viscosity is less likely to penetrate into, and flush out the smaller crevices where microorganisms lurk and flourish upon food debris. Mucin may also encourage decay by forming a tenacious coating upon a sheltered surface of the tooth and thus serve to protect underlying bacteria from the action of the saliva.

The investigations of Pincus suggest that decalcification of the teeth by acid is of less importance than an attack upon the organic substance of the enamel by proteolytic bacteria. Such a process has been suggested earlier by Hines. Pincus claims that digestion of the

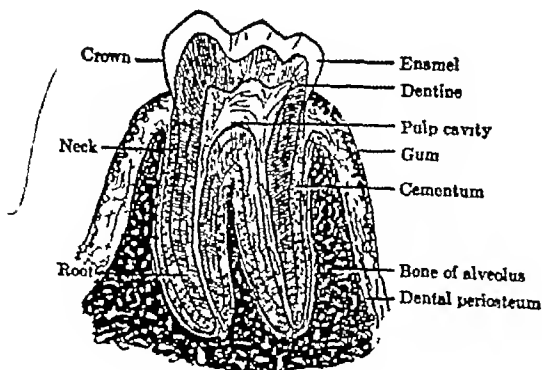


FIG 374 Section through a molar tooth

¹ This action is analogous to the effect of sugar in favoring the penetration of toxins.

protein material permits the enamel prisms to fall apart, and this occurs apparently without a preliminary production of acid. Sulphuric acid liberated by the hydrolysis of chondroitin sulphate, according to Pincus causes, secondarily, decalcification of the enamel. Atkinson and Matthews agree that the primary point of attack is upon the organic matrix of the enamel, but that decalcification of both enamel and dentine is produced by the liberation of aspartic and glutamic acids. They found some 18 amino acids present in normal dentine, including aspartic and glutamic acids. These two amino acids were found in the free state in carious dentine but not in healthy dentine. The well-known brown discoloration of carious teeth is due to a melanin which is thought to be formed by the oxidation of tyrosine by proteolytic bacteria.

The researches of M. Mellanby on the teeth of dogs suggest that, in some cases at least, the structure of the teeth themselves—maldevelopment, or an inherently poor tooth structure—or dietary deficiencies (especially of vitamin D) may play a part in the occurrence of the disease. Contrary to the general belief, caries in wild animals is not so very uncommon. It starts, with few exceptions, in the dentine or cementum of the root, where this has become bare, and not in the enamel. The decay is most likely to occur in situations where food tends to pack and stagnate. In domestic animals and animals in captivity, caries is much commoner and the first signs of decay appear in the enamel, as in man. A survey of some 280 wild animals belonging to 60 different species were made by Sprawson, and the tooth enamel compared with that in human teeth. The structure of the animal enamel was found to be less dense, less fully calcified and, on the average, of poorer quality than that of human enamel, a finding which, taken in conjunction with the less common occurrence of caries in wild animals, suggests a dietary factor in the production of the disease in man.

The beneficial effects of small amounts of fluorine upon tooth structure has received no small amount of attention since Dean and his associates pointed out the low incidence of caries in districts where there was evidence of high concentrations (over 1 part per million) of fluorine in the drinking water. Excessively high concentrations of fluorine in the drinking water cause mottling of the enamel. Fluorine has also been shown to inhibit acid production by the *Lactobacillus acidophilus* and in a concentration of 0.5 to 1 part per million to be protective against dental caries. Another possible action of fluoride is to reduce the solubility of the enamel by acid. Stannous fluoride is more effective as a preventive than the sodium salt. From the dental examination of some 7,000 children in over 20 American cities it was found that the incidence of caries was inversely related to the concentration of fluorine in the drinking water. Mottling of the enamel does not occur in concentrations below 1 part in a million.

A curious condition, known as *dermeus*, occurring in domestic animals and man is seen in North Africa, and is attributed to a high concentration of calcium fluoride in the drinking water. It is characterized by dystrophy of the second dentition. Kilborn has also reported the endemic occurrence of a severe arthritis and spondylitis in a region of Southwest China which apparently is induced by the drinking of water rich in fluoride.

It is quite evident that the problem of dental caries is not a simple one. Several factors are involved in its production and much work must be done before the details of their actions are known and the relative value of each can be appraised. But from the point of view of prevention the most valuable measures are the curtailment of sugar in the diet and the addition of fluoride to the drinking water in those communities in which the latter has a low concentration of this element.

GASTRIC DIGESTION

STRUCTURE OF THE GASTRIC GLANDS

The secretory cells of the stomach are arranged to form innumerable small individual tubular units, the *gastric glands*. The glands are scattered profusely throughout the mucosa, each has a depth of from 0.5 to 3.5 mm and an outside diameter of 0.1 mm. The lumen of the tubule is not more than 10 micra in caliber. A tubule empties by a single mouth into the bottom of a tiny pit—the *gastric foveola* or *cripi*. In some cases a single foveola receives the secretion of two or more glands. The gastric foveolae dot the mucosal surface in immense numbers, they are lined by columnar epithelium, which has a similar character in all parts of the stomach, and is continuous, on the one hand, with the general epithelial covering of the gastric mucosa, and, on the other, with the cells lining the gland tubules. In the latter situation the characters of the cells undergo a change, both morphological and functional, and acquire secretory powers of a special nature.

Each gland may appear as a single tubular structure which lies more or less perpendicular to the surface of the mucosa. More usually, its deeper portion is branched or shows an ill-marked bifurcation. The cellular elements of the glands of the fundus and of the greater part of the body of the stomach differ both in minute structure and in the nature of their secretions from those of the pyloric region, as well as from those of the mucosa in the immediate neighborhood of the cardia.

(1) *The glands of the fundus and body of the mammalian stomach*

The gland tubule is divisible into a narrow and short superficial part—the *neck*, and a much thicker and longer portion—the *body*, which reaches nearly to the muscularis mucosae. Its blind extremity is club-shaped, indented or branched. The total number of glands has been estimated at 35,000,000. The cells which constitute the walls of the tubules are of three types: (a) *chief cells of the neck or mucous neck cells*, (b) *chief cells of the body or zymogenic cells*, and (c) *parietal or border cells*.

The *chief cells of the neck or mucous neck cells* together with the *chief cells of the body* form a continuous lining for the tubule, and upon a superficial examination resemble one another very closely. The two types of

cell are, however, quite different both in minute structure and function. Their morphological differences can be readily brought out by special staining methods. The chief cells of the neck contain granules of mucinogen but no zymogen material. Their secretion is, in consequence, entirely mucous in character. These cells are identical with those lining the entire length of the pyloric glands (see fig. 38.1).

The *chief cells of the body* or *zymogenic cells*, on the other hand, contain zymogen but no mucinogen granules. The former are composed of pepsinogen, the precursor of *pepsin*. There is no doubt that the chief cells furnish the enzymes of the juice, after secretion has been evoked by vagal stimulation the granules become markedly reduced in number, they accumulate again after a period of rest in the same way as has already been described for the zymogen granules of the salivary cells.

The *parietal or border cells* are found throughout the entire length of the tubule, but are more numerous in the region of the neck. They do not, however, form a continuous layer, but are scattered here and there along the tubule and lie, not against the lumen, but between the chief cells and the basement membrane. They bulge the membrane outwards and in this way give an uneven, nodular or wavy contour to the gland. These cells can be distinguished from their neighbors not only by their location but by their staining reactions. It is now generally agreed that the parietal cells secrete the hydrochloric acid and most of the water of the gastric juice, and on this account are also known as the *oxyntic* or *acid secreting cells*. Among the several observations which point to this conclusion is the direct correlation between the number of parietal cells and the titratable acidity and total chlorine contents of different regions of the gastric mucosa during secretion. Though separated from the lumen of the tubule by the chief or zymogenic cells, the parietal cells transmit their secretion into the lumen through delicate canaliculi lying between the chief cells. These canaliculi are extensions of an exquisitely fine canal system lying within the cytoplasm of the parietal cells and possess definite walls. Coarse transparent granules are seen in the protoplasm of the cells during rest, depletion of this granular material occurs during activity.

(2) *The pyloric, esophageal and cardiac glands*

The pyloric glands are shorter but run a more tortuous course. Parietal cells are absent, and the tubules are lined entirely by cells of the same type as the mucous neck cells of the fundic glands. The

pyloric glands, in consequence, are capable of forming mucus but not of supplying pepsin or hydrochloric acid to the gastric juice. Their secretion is alkaline in reaction. The glands in the immediate neighborhood of the cardiac orifice—cardiac glands—are also of the purely mucous type and secrete an alkaline juice. Glands almost identical in character are present in the esophagus—at the level of the cricoid cartilage and at its lower end just above the cardia.

THE GASTRIC JUICE, COMPOSITION, ORIGINS AND DIGESTIVE ACTIONS OF ITS CONSTITUENTS

(SEE ALSO INTRINSIC FACTOR, P. 85)

COMPOSITION

The gastric juice is a mixture of the secretory products—*pepsin*, *mucin*, and a watery solution of *hydrochloric acid*—of the three types of cells composing the gastric glands described above, it also contains a variable quantity of mucus secreted by the surface epithelium. The composition of fasting human gastric juice as obtained through a fistula is given in the following table.

Acidity	Free HCl, 0.40 to 0.50 per cent
	Total acidity, 0.45 to 0.60 per cent
	pH, 0.90 to 1.0
Solids	Organic, including mucin and the various ferments, 0.42 to 0.46 per cent
	Inorganic, 0.13 to 0.14 per cent
Specific gravity, 1.002 to 1.004	

A proteolytic enzyme, called *uropepsin*, which is found in urine, apparently has its source in the gastric glands, for it disappears from the urine after gastrectomy or removal of the gastric mucosa. Its concentration in the urine is increased by high protein feeding but not by pepsin introduced into the stomach. It is thought therefore to be derived from pepsinogen which is absorbed directly into the blood stream from the glands.

The stomach secretes extraneous substances e.g., ether, sulfonamide, drugs, iodides and bromides, neutral red, etc.

The main digestive action of the gastric juice is exerted upon protein, which is split into smaller groups of amino acids by *pepsin* and *hydrochloric acid*. It also contains *rennin* which curdles milk and a very weak lipolytic ferment—*gastric lipase*.

The gastric juice—a combined secretion of several types of cells—consists of an alkaline and an acid component, which vary inversely in their

proportions with the volume of the secretion, the alkaline component falling and the acid component rising with increased rate of secretion and vice versa. The alkaline component, which is secreted by cells other than the parietal cells, contains pepsin, mucin, neutral chloride, sodium, bicarbonate, potassium and calcium, the acid component is the solution of hydrochloric acid secreted by the parietal cells.

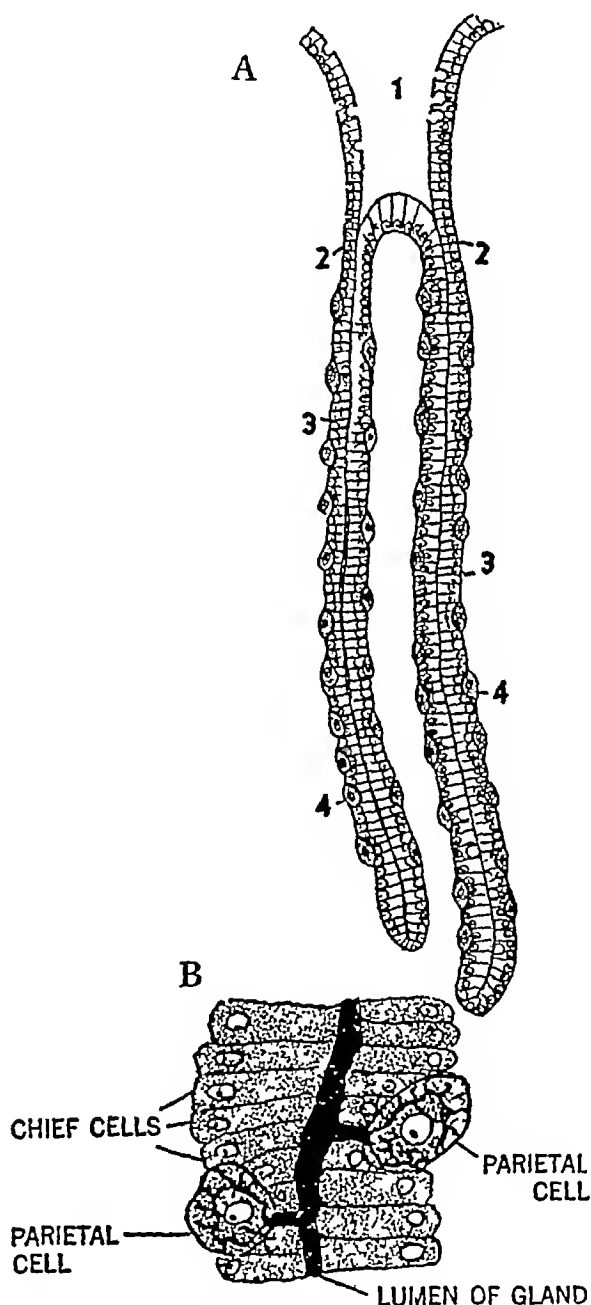


FIG. 38.1 A, gland from the fundus of the stomach 1, pit or foveola on the mucous surface, 2, neck of gland containing mucin-secreting cells, 3, chief cells, 4, parietal cells B, enlarged drawing showing intercellular and intracellular canaliculi

HYDROCHLORIC ACID IN THE GASTRIC JUICE

Prout in the early part of the last century (1824) first demonstrated the presence of hydrochloric acid in gastric juice. The only other instances of a strong mineral acid being formed by living processes is the production of sulphuric acid, probably for defense purposes, by a certain snail (*Dolium galea*), and the presence of a fluid with a pH as low as 1.0 in the vacuoles of some unicellular organisms. It has been already mentioned that the parietal cells of the fundic glands are responsible for the production of the acid and most of the water of the gastric juice.

The secretion of the parietal cells cannot be collected separately from the other constituents of the gastric juice and, therefore, cannot be subjected to direct analysis. Its composition can be arrived at only by inference. Hollander analyzed canine gastric juice collected from a stomach pouch and has inferred the composition of pure parietal secretion from the fact that, with increasing acidity, the other constituents decline. The acidity when plotted against the neutral chloride gives a straight line expressing an inverse relationship between them. When extrapolated this line comes to a point representing a zero value for neutral chloride and an acidity of around 170 m eq per liter (figs 38.2, 38.3, 38.4). Changes in total chloride run parallel with those for total acidity, though the former occur at a much slower rate. From this it is concluded that pure parietal secretion is a slightly hypertonic solution fluid containing approximately 170 m eq HCl per liter and free from neutral chloride, its pH is around 0.90. Pavlov found that with rising rate of secretion

the concentration of acid in the gastric juice rose and concluded that the acidity of the gastric juice as it left the cells remained constant (regardless of the rate or volume of secretion) at a concentration of about 0.5 per cent. Pavlov's belief in the constancy of the acid concentration as secreted by the glands is supported in the main by Hollander's experimental results and this view is now widely held.

The contents of the human stomach after the usual test meal have a total acidity of from 5 to 40 clinical units (p. 515) or from 0.02 to 0.15 per cent of total acid. The gastric juice secreted into the empty stomach after the more powerful stimulus of histamine injection or in response to psychic stimulation has a total acidity of from 80 to 140 clinical units and a pH of from 1.0 to 1.2. At the height of digestion of a mixed meal containing meat, the acidity of the gastric contents may approach that of juice secreted under the influence of histamine. The relatively low acidity of the human gastric contents after a test meal is due to the latter's relatively mild stimulating effect upon the glands as well as to the dilution and neutralization of the juice by saliva, mucus and other constituents of the meal itself. The analyses of histamine juice, as noted above, and of pure juice collected from a fistula indicate, however, that human gastric juice as it is secreted by the glands has probably an acidity equal to that of canine juice.

There is a relationship (in dogs and human subjects) between the acid concentration in the gastric juice and the erythrocyte count (Apperly and Cary). In severe anemia due to hemorrhage free acid disappears from the gastric juice, as blood regeneration occurs acid reappears, its concentration increasing to a maximum when the red cell count reaches the normal value. If the erythrocyte count rises above the normal level (e.g., in polycythemia) the acid concentration steadily falls again to low values.

Theories concerning the origin of the HCl and site of its production

Claude Bernard was one of the first to investigate the question of the origin of the acid. He injected potassium ferrocyanide and lactate of iron into an animal's veins. These substances give the Prussian blue reaction in the presence of free mineral acid. The mucosa of the stomach was turned blue after the injection, but no discoloration was observed in the gland cells or in the lumina of the glands. He concluded, therefore, that the

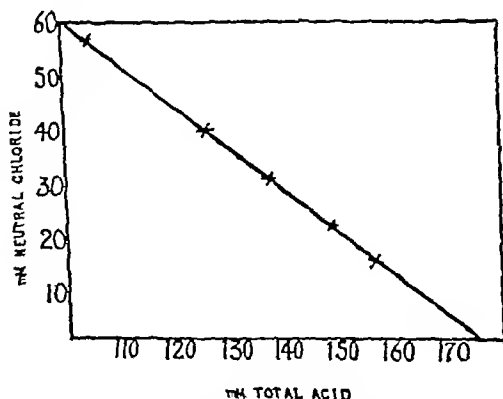


FIG 38.2 Showing neutral chloride as a function of acidity (After Hollander)

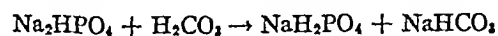
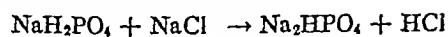
glands secreted some unknown product, which, upon reaching the surface of the mucosa, gave rise to the acid

Similar experiments were carried out by Fitzgerald, and later, vital dyes which changed color at definite pH values were employed by Harvey and Bensley. The results of these two groups of investigators were not in agreement, the former claiming that the acid was formed in the cytoplasm or within the fine intracellular canaliculi. Harvey and Bensley suggested that NH_4Cl was secreted into the tubules and through selective re-absorption of ammonia, HCl was formed. Subsequently, however, Ivy and Dawson by means of a Pavlov pouch and the use of vital dye indicators demonstrated that during secretion the cells have a pH ranging somewhere between 6.8 and 3.0. A more precise value was not determined, but the result suggested that the site of acid formation was intracellular. Hollander believes that the most probable site is the cytoplasmic surface of the intracellular canicular wall or within the wall itself. This site of origin was first suggested by Collip.

The ultimate source of the chloride is undoubtedly the sodium chloride of the blood. Kahn showed many years ago that, if an animal be fed for some weeks upon meat, deprived of its chlorides by prolonged boiling in distilled water, acid-free gastric juice was secreted. Also the continued loss of gastric juice from the body results in a marked fall in blood chloride. If bromide is fed to an animal which has been deprived of salt for some time, hydrobromic acid appears in the gastric juice. The chloride ion can also to some extent be replaced by iodide.

Several theories of the chemical processes concerned in the formation of hydrochloric acid from the blood chloride have been advanced,

Many believed reactions represented by the following equations were responsible



The acid formed in the first reaction would have to be removed as quickly as it was formed and fresh acid phosphate (NaH_2PO_4) furnished by the reaction shown in the second equation. Maly could offer no explanation to account for the separation of the products of these reactions, that is, the passage of HCl into the tubules of the gland and the bicarbonate into the blood stream. Collip by microchemical methods obtained important

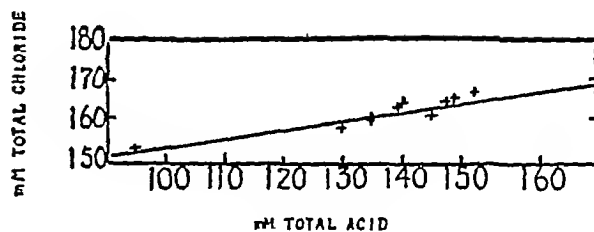


FIG 38.3 Total chloride as a function of acidity (After Hollander)

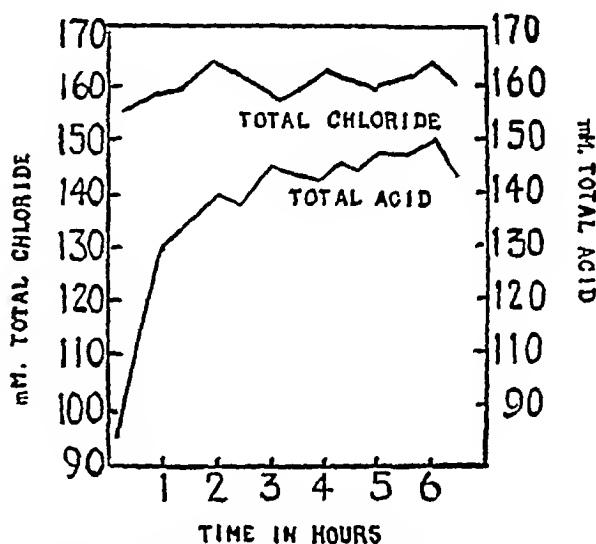


FIG 38.4 Total chloride as a function of time (After Hollander, redrawn)

evidence in support of this conception. He states that the parietal cells are exceptionally rich in phosphates and probably in carbonates, but during rest chlorides are absent. When, however, the glands become active a remarkable increase in the cell chlorides occurs. It can also be shown that during secretion the chloride content of the arterial blood supplying the gastric mucosa decreases and the bicarbonate of the venous blood increases.

The relationship between the CO_2 content of the blood and the secretion of acid in the gastric juice has been demonstrated by several workers. Hyperventilation (blowing off CO_2) reduces, and breathing CO_2 rich mixture increases, the acid secretion. Apperly and Crabtree found that the gastric acidity in man varied directly with the bicarbonate content of the plasma. Browne and Vineberg have demonstrated that the gastric secretion induced by vagal stimulation in dogs is inhibited when the CO_2 content of the plasma falls below 30 volumes per cent as a result of hyperventilation but is restored again when the CO_2 percentage of the inspired air is raised.

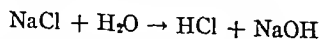
Davenport and Fisher found a high concentration of carbonic anhydrase in the parietal cells of the gastric mucosa, and pointed out that acid secretion is proportional to the rate of formation of H_2CO_3 from CO_2 and

H_2O It was proposed that this was dependent upon the action of the enzyme and that the H -ions of the parietal fluid were derived from carbonic acid, the bicarbonate ion passing into the blood stream. But it was found that thiocyanate which inhibits gastric secretion had not an inhibitory effect upon the action of the enzyme to a corresponding degree, while sulfanilamide and thiophene 2 sulfonamide, which have profoundly depressing effects upon the action of carbonic anhydrase do not inhibit gastric secretion. The theory was then withdrawn by its authors. However, synthetic sulfonamide derivatives¹ have since been obtained which inhibit the action of the enzyme, as well as, the gastric secretory response to histamine. The probable role of carbonic anhydrase is now thought to be the replenishment of the store of H ions in the parietal cell through its well known action in catalyzing the hydration of CO_2 to H_2CO_3 .

Other theories have been proposed suggesting that H ions are derived from an organic acid or chloride, which is synthesized by the parietal cells and secreted into the canaliculus, then by hydrolysis, or double decomposition with $NaCl$, or by selective reabsorption, HCl is produced. Bull and Gray, for example, postulate the secretion of an organic acid (possibly pyruvic or lactic) into the intracellular canaliculus at its blind end. Beyond this, it is supposed, the canalicular membrane is impermeable to cations but permits the exchange of Cl^- (from $NaCl$) within the cell with anions of the organic acid. In order for a positive concentration gradient to be sustained, the H ions of the latter then combine with Cl to form HCl , and thus to permit the rapid diffusion of the organic anion across the membrane from the canalicular lumen, the latter must be rapidly destroyed. This is accomplished supposedly by decarboxylation. The released carbon dioxide is then hydrated to H_2CO_3 (the conversion catalyzed probably by carbonic anhydrase) which reacting with the released alkali passes into the blood stream as bicarbonate.

Hoerr and Bensley postulate the secretion into the canalicular lumen of a protein hydrochloride which is progressively diluted by non-parietal secretion as it passes along the tubular lumen to the foveola, here, supposedly, it is hydrolysed and HCl set free.

The simplest, as well as the most acceptable theory, is a physicochemical one proposed by Hollander, based upon hydrolysis of $NaCl$ which enters the cell from the interstitial fluid at the intracellular canalicular wall, and separation of the hydrolytic products through the irreciprocal permeability of the membrane. Thus



¹ One of these is 2 acetylamine 1,3,4-thiodiazole 5-sulfanilamide (No. 6063)

H ions, water and Cl ions can pass across the intracanalicular membrane from the cytoplasm, but not in the reverse direction. A nearly isotonic solution of HCl is thus secreted and moved along the canaliculus to the tubular lumen. The $NaOH$ is buffered within the cell and passes with water as bicarbonate or the alkaline phosphate into the tissue fluids and hence into the circulation. Such a process, namely, the *membrane hydrolysis* of $NaCl$, with the formation of HCl and $NaOH$, the latter being retained and buffered, has been described as occurring in purely physicochemical systems, which lends color to the suggestion that it may occur in the parietal cell.

The foregoing account by no means exhausts all the theories which have been proposed in respect to the formation of gastric acid (see Conway). But in spite of the many ingenious attempts which have been made to explain the origin of the H ions, the problem is still unsolved. Yet it can be said that however they are derived, their secretion results in secondary interchanges which are well authenticated. The secretion of H^+ entails a corresponding passage of Cl^- across the canalicular membrane, in order that electrical neutrality shall be maintained. The venous blood leaving the gastric mucosa shows a fall in chloride and a rise in bicarbonate. A *reverse chloride shift* then occurs between the plasma and red cells, HCO_3^- entering the cells and Cl^- leaving the cells for the plasma.

The work of the parietal cells The work performed by the parietal cells has been divided into three categories by Hollander, namely, *chemical*, performed in the reactions leading to the production of HCl at the canalicular membrane, *electrical*, in the establishment of an electrical potential on the two sides of the membrane required for the separation of the HCl from the other reactants, and *mechanical*, required for the movement of the secretion along the fine canaliculi. There have been several attempts to calculate the energy expended in the performance of the glandular work. The results of different workers vary from 0.9 to 1.5 calories per liter of parietal fluid secreted. But Hollander points out that these calculations were directed toward the estimation of osmotic work, and that the parietal cell probably does not perform work of this kind. It does not, as does the cell of the renal tubule (p. 453), separate a fluid (reabsorbed fluid) from one of higher concentration in osmotically active substances, but secretes a solution isotonic with the blood plasma. The performance of mechanical work is evident from the fact that distension of the canaliculi can be observed during secretion. The manner in which the distending and propulsive force is effected is quite unknown.

The relation of gastric secretion to the acid-base-balance

A reduction in the excretion of acid by the kidney—the so-called *alkaline tide of the urine*—occurs after a meal. This was attributed to the loss of HCl in the gastric juice and the secondary rise in bicarbonate. Higgins, and later, Dodds and Bennett found that a rise of from 2 to 5 mm in the CO₂ tension of the alveolar air occurred half an hour or so after meals, a corresponding fall below the normal level was reported to follow about one and a half hours later. These changes have been looked upon as compensatory reactions brought about by the loss of acid in the gastric juice, and of alkali in the pancreatic juice, respectively, which served to maintain the normal acid-base balance of the body fluids. Though the reduced excretion of acid by the kidney and the rise in the CO₂ tension of the alveolar air and plasma do actually occur during normal gastric digestion, recent work has seriously questioned the generally accepted explanation that the phenomena are due to the secretion of HCl (see p. 462).

When, on the other hand, abnormally large quantities of gastric juice are lost from the body, as in pyloric obstruction with persistent vomiting, or in intestinal obstruction, a profound effect upon the acid-base balance is produced. A fall in blood chloride leading to a state of compensated or uncompensated alkalosis results, tetany may occur. Dragstedt and Ellis have also shown that, in animals, drainage of the gastric juice to the exterior through a fistula results in a reduction of 50 per cent in the blood chloride, alkalosis, marked dehydration and a rise in the nonprotein nitrogen of the blood. Severe depression occurred ending in death.

THE ACTIONS OF THE GASTRIC ENZYMES, MUCIN

PEPSIN is derived from the chief cells of the fundic glands, evidence for this has already been cited on p. 496. The zymogen granules in the cytoplasm of these cells are believed to represent the mother substance of the active ferment. In this form the enzyme is inactive and is usually referred to as pepsinogen. The enzyme becomes active only in acid media, the optimum pH value for its action is about 1.5 (Sorensen). The optimum pH varies, however, in accordance with the particular protein which is being digested. Thus, the optimum pH for the digestion of casein is 1.8, for the digestion of gelatin 2.2. At a pH of 5.0 the action of pepsin is almost abolished. The digestive action of pepsin is confined to protein which it splits into proteoses and peptones. Pepsin has been crystallized, but it is not a single substance, it is composed of at least three proteins only two of which are proteolytic. One of the latter has been recrystal-

lized and appears to be a pure homogeneous substance. The crystals of pepsin are doubly refracting hexahedra.

The first stage in gastric digestion is the action of the acid upon protein to form acid metaprotein. This product is insoluble in water and neutral solutions but soluble in the acid secretion of the stomach. The acid metaprotein is then acted upon by pepsin. The protein molecule is constructed of many amino-acids (p. 622), the greater proportion of which, it is generally believed, are linked together by their amino (NH₂) and carboxyl (COOH) groups. This is spoken of as the *peptide linkage*. The function of the different proteolytic enzymes of the digestive juices (gastric, pancreatic and intestinal) is to break the protein molecule into smaller and smaller fragments containing successively fewer numbers of amino-acids. A molecule of water is taken up (hydrolysis) as the first step in the process and cleavage occurs at the CO—NH junction.² In the final stage of protein disintegration the molecule is completely disrupted and its individual "building stones"—the amino-acids—separated from one another. This ultimate stage must be reached before protein material can be made use of by the body. Gastric digestion of protein does not go beyond the peptone stage. The action of pepsin (in vitro) is reversible. Wasteneys and Borsook have demonstrated the synthesis of protein by pepsin from a concentrated peptic digest of albumin. The optimum hydrogen ion concentration for the synthetic action is at pH 4.0.

GASTRIC RENNIN (OR RENNET) This, the milk-curdling ferment, is generally believed to be a product also of the chief cells of the fundic glands. It is found in the 4th stomach of the calf but appears to be absent from the human stomach—of both adults and infants—the clotting and digestion of milk being affected by pepsin. The optimum pH for the action of rennin is between 6 and 6.5 and it is quite inactive at the pH of the gastric contents of the normal adult.

The clotting of milk by rennin shows a certain resemblance in several of its chemical and physical features to the clotting of blood (ch. 12). The resem-

² Though this type of cleavage is unquestionably effected by the pancreatic and intestinal proteolytic enzymes, some have doubted that pepsin is capable of breaking the peptide linkage and have suggested that it attacks some other type of linkage in the protein molecule (see diketopiperazine linkage, ch. 46). The evidence, however, is strongly in favor of the view that the peptide linkages are broken by peptic action.

blance, however, is purely superficial, fundamentally the two processes are quite different. In each instance a soluble protein is rendered insoluble as the result of the activity of a ferment, and in order that this shall occur it is necessary that calcium salts be present in ionized form. In both processes a clear incoagulable fluid, whey in the case of milk, serum in the case of blood, can be expressed or separates spontaneously when the coagulum is allowed to stand. Oxalates and citrates by precipitating the calcium prevent the clotting of either of these fluids. Gastric rennin when added to milk kept at body temperature causes a change in the soluble *casein* (caseinogen) splitting it into a *proteose-like* substance (*whey protein*) which remains in solution, and *paracasein*. The latter is soluble itself but upon combination with calcium forms an insoluble compound, *calcium paracasein* which is thrown down to form the clot or curd. The protein (paracasein) of the curd subsequently undergoes peptic digestion in the usual manner. The softer and more finely flocculent the character of the curd that is formed, the more readily is it attacked by the proteolytic ferments. According to Pavlov the admixture of mucus with the milk renders the resulting curd softer in consistency.

Gastric lipase is a weak fat-splitting enzyme. It differs in an essential manner from the lipases of the pancreatic and intestinal juices in that it acts in an acid medium, and is reduced in activity and finally destroyed by alkali. Its pH optimum ranges between 4 and 5. At a pH of 2.5 its action ceases. It is assumed that gastric lipase is also secreted by the chief cells of the fundic glands, though proof of this is lacking.

Owing to its very weak action gastric lipase is of little practical importance, in adults at any rate. It is soon rendered inactive as the acidity of the gastric contents rises during digestion, and any action which it may have, is exerted only upon fats in a state of very fine emulsion, e.g., milk, yolk of egg, etc. On this account it may possibly be of greater importance in infants in whose stomachs it is found at, or shortly after birth, and whose gastric contents have a pH more favorable for its action. The whey of mother's milk is said to enhance the action of gastric lipase in infants. Nevertheless, it is questionable whether this enzyme even in infants possesses any real digestive value.

According to some observers digestion of fat in the stomach may be accomplished by pancreatic lipase that has regurgitated through the pylorus. But since pancreatic lipase acts best in a decidedly alkaline medium, its optimum pH being 8, and very weakly in an acid medium, its action must be slight so long as the gastric juice possesses its normal acidity. If, however,

this is depressed or absent the gastric digestion of fat may be effected to a considerable extent by pancreatic lipase. It is to be remembered in this connection that a meal of fat depresses gastric acidity. Fat also increases duodenal regurgitation which by aiding neutralization may bring about the reaction suitable for the activity of pancreatic lipase.

Less important digestive enzymes of gastric juice or of gastric tissue are the protease, *cathepsin*, the carbohydrate splitting enzyme, *lysozyme*, and a *gelatinase* which acts specifically upon *gelatin*. The enzyme in gastric tissue known as the intrinsic factor is dealt with in chapter 9.

Gastric mucin. Mucin is a glycoprotein. It is secreted by the cells of the pyloric and cardiac glands and by the mucous neck cells of the fundic glands. It is therefore a constituent of pure gastric juice (dissolved mucin). It is also a constituent of the mucous secretion of the ordinary epithelial cells of the gastric mucosa (surface epithelium mucus). Ivy found that the secretion collected from pouches fashioned from the pyloric part of the stomach of dogs was mucoid, viscous, tenacious and transparent with a pH of 7.0 to 7.5. But gastric mucus varies considerably in consistency, being sometimes a clear fluid of low viscosity and at other times thick, viscous and even jelly-like.

Gastric mucin has a high acid combining power. It coats the interior of the stomach, and by reducing the free hydrochloric acid serves to protect the mucosa from the action of the gastric juice. Quite apart from its power to lower the acidity of the gastric juice Babkin and Komarov find that mucin inhibits peptic activity. This anti-peptic action of mucin is due to its constituent, mucicetin-sulphuric acid. Its physical properties also serve to lubricate and protect the gastric mucosa from mechanical injury. The protective properties of gastric mucin have been put to practical use by Fogelson and by Ivy and Kim, in the treatment of gastric and duodenal ulcer. The mucin is prepared from hog's stomach. One gram of such a preparation combines with 15 cc. of a 0.5 per cent hydrochloric acid and Fogelson found that about 60 grams was more than sufficient to neutralize the acid secreted in response to the injection of 1 mg. of histamine. Also when $\frac{1}{4}$ ounce of mucin together with 1 pound of meat was given to a dog, free acid did not appear in the stomach until the lapse of from 5 to 7 hours.

Mucous secretion from the surface epithelium is stimulated by various chemicals applied to the surface of the gastric mucosa, e.g., alcohol, ether,

clove oil, phenol red, etc. Vagal stimulation evokes a secretion of mucin from the mucin-secreting cells of the tubules

THE CHYME

The changes in the food brought about by the various chemical and physical processes that have been described as constituting gastric digestion may be briefly summarized. Much of the *protein* of the food is reduced to simpler materials which are soluble in the fluids of the stomach. The *fats* undergo some degree of emulsification. Other materials, such as the *sugars* taken preformed with the food or those derived from the partial digestion of the starches, pass readily into solution. The remainder of the *starch* is in part rendered soluble (soluble starch and dextrins) by salivary action and in part reduced to a fine state of mechanical division. The gastric movements (p. 564) cause these various elements, which differ widely in their chemical and physical natures, to become thoroughly mixed with one another and with the gastric juice until finally the food assumes a semi-fluid, more or less homogeneous, creamy or gruel-like consistency. This material, now known as the *chyme* (Gk. *chymos*—juice) is definitely acid in reaction, it passes through the pyloric opening, not all at once but from time to time as it is formed. The food having reached this stage is suitably prepared for further digestion by the intestinal juices

THE RELATIVE IMPORTANCE OF CHEMICAL AND MECHANICAL FACTORS IN GASTRIC DIGESTION

Borelli and the mechanical school of physiologists of the seventeenth century saw digestion simply as a process in which crushing, grinding and mixing of the food, and the expression of its nutritive juices were the prime features. Chemical processes were scarcely considered. The second stomach or gizzard of the bird in which resistant objects, even glass beads, are crushed was pointed to in support of these mechanical conceptions of gastric digestion. Not until the experiments of Réaumur (1683–1757) and Spallanzani (1729–1799) was the chemical character of gastric digestion proved beyond doubt and fully appreciated. They showed by a crucial experiment—the digestion of meat by gastric juice in a test-tube—that gastric digestion was a chemical process, and not simply a mechanical disintegration of the food. Ever since, more stress, perhaps, has been laid upon

the chemical than upon the mechanical side of digestion. It is probable, however, that the latter, insofar as gastric disorders are concerned, is in reality of just as great or even of greater importance than the chemical factor. Subjects in whom the gastric juice is entirely absent usually suffer no apparent difficulty in the digestion of food and there may be no gastric symptoms whatsoever. The larger part of the stomach may be removed and, provided that the consistency of the food is made suitable, little digestive inconvenience results. After complete removal of the human stomach, protein and fat digestion have been found to be perfectly normal, but there may be depression of intestinal motility as a result of the unavoidable severance of the vagus nerves. Purely mechanical factors, on the contrary, very frequently give rise to digestive disturbances, and it is probably true that the *immediate* factors responsible for the production of the various symptoms of gastric disease are always mechanical in nature. Disturbances of gastric motility, for example, may cause discomfort, pain and other dyspeptic symptoms. The stomach contents may be retarded in their passage into the duodenum as a result of spasm of the pylorus or of an organic blockage. On the other hand, a too rapid evacuation of the stomach may, by permitting food to enter the duodenum in an improperly prepared state, induce ill effects. The chemical processes of gastric digestion, therefore, important as they are, must not be allowed to overshadow in our minds the mechanical factors. A prime function of the stomach is to break up the food, to add fluid to it and reduce the entire mass to a semifluid consistency, in which state it can be more readily acted upon by the digestive enzymes of the intestine. The movements of the stomach will be discussed in chapter 42.

THE ANTISEPTIC ACTION OF GASTRIC JUICE By virtue of its high acidity normal gastric juice has an important bactericidal action. Streptococci, staphylococci and *B. coli* are destroyed, the contents of the duodenum in health being virtually sterile. In permanent or temporary achlorhydria, on the contrary, the duodenum is very quickly invaded with microorganisms from the colon. Following perforation of a gastric or duodenal ulcer, for example, the gastric secretions are suppressed as a result of the shock occasioned by the accident, and peritonitis supervenes after sufficient time has elapsed for the duodenal fluids to become infected in this way. The importance of early operation is self evident.

The effects of gastrectomy upon certain non digestive

functions Since the stomach plays an important part in erythropoiesis (p 109), it might be expected that anemia of the pernicious type would follow its removal but, as a matter of fact, pernicious anemia is rarely seen after total gastrectomy either in man or in animals. The microcytic hypochromic type, on the other hand, is not uncommonly a result of and is most probably due to defective absorption of iron caused by the lack of hydrochloric acid (p 76). It responds to iron administration. Ivy and his associates have reported rarefaction of the bones in puppies leading to gross deformities of the limbs following gastrectomy. Removal of the stomach in growing monkeys caused arrested growth and a condition of the bones resembling osteitis fibrosa cystica (p 858), together with hypocalcemia and enlargement of the parathyroid glands. The condition is the result apparently of impaired calcium absorption due in turn to the absence of hydrochloric acid (p 861), it can be prevented to a large extent or partially corrected by adding soluble calcium salts to the diet. As well as the one just mentioned, Ivy lists three other causes of the osteoporosis, (a) the absence of the reservoir function of the stomach, imperfectly digested

food being passed rapidly through the upper part of the intestine from where the greater part of the calcium is normally absorbed, (b) the postprandial acidosis resulting from the secretion of alkaline intestinal juices in the absence of the acid gastric secretion and (c) encroachment of the red marrow (due to the anemia) upon the osseous tissue.

In man, a not uncommon result of partial gastrectomy is a temporary hyperglycemia and glycosuria after a meal, followed by a longer period of hypoglycemia. The hyperglycemia is attributed to the rapid passage of food into the intestine which thus results in an increased rate of absorption. Barnes suggests that the recurrence of hyperglycemia after each meal causes the subject to become hypersensitive to his own insulin, thus hypoglycemia results. Since the hypoglycemic response does not develop at once after operation, but is delayed for weeks or months, increased output of insulin is not thought to be a likely explanation. The failure of adrenaline to be secreted as the blood sugar is falling has also been suggested as an explanation, but signs of adrenaline secretion are usually displayed by these patients.

CHAPTER 39

GASTRIC DIGESTION (*Continued*)

THE SECRETION OF GASTRIC JUICE

THE INNERVATION OF THE GASTRIC GLANDS

Stimulation of the vagus nerve causes the secretion of juice high in peptic power and strongly acid. This action is mediated through acetylcholine. According to Alley and Babkin the glands of the lesser curvature respond more readily to vagal stimulation than do those of the greater curvature, and secrete a stronger juice. This fact may be of significance in the development of gastric ulcer (p 520). Baxter has shown that stimulation of the sympathetic causes the secretion, mainly from the pyloric glands, of an alkaline mucoid juice which is very low in peptic power, this secretion is unaffected by atropine but is annulled by ergotamine (which paralyzes motor and secretory sympathetic fibers). Vineberg's results indicate that the vagus controls the secretion of mucin by the mucous neck cells and the surface epithelium of the gastric mucosa. The influence exerted by the sympathetic upon the peptic and oxyntic cells of the gastric body is not definitely known. According to some observers its effect upon these elements is inhibitory.

THE QUESTION OF THE CONTINUOUS OR INTER-DIGESTIVE SECRETION OF GASTRIC JUICE

Pavlov found that the secretion of gastric juice in dogs was intermittent, beyond the secretion of some alkaline mucus the gastric glands in the absence of food or psychic influences (see below) remained at rest. This conclusion has been confirmed by the recent experiments of Babkin. But in man, juice is secreted continuously in fairly large amounts, though this fact does not necessarily indicate a fundamental difference between the activities of the glands of the human and canine stomachs. The shorter intervals between meals, and psychic influences which are impossible to eliminate in the case of the human subject, are probably responsible for the continuous and apparently spontaneous secretion. This secretion is increased during sleep.

THE PHASES OF GASTRIC SECRETION

It has been known for many years that it is not a necessary condition for the excitation of the gastric glands that food shall enter the stomach. The mere presence of food in the mouth calls forth an abundant secretion of gastric juice. The flow of juice that occurs in this instance is due to a nervous reflex and is termed the *psychic* or *cephalic phase of gastric secretion*. It is also well known that the glands are stimulated by the presence of food in the stomach, even after all nerves connecting the organ with the central nervous system have been severed. This secretion of gastric juice which results from influences arising within the stomach itself is referred to as the *gastric phase*. Secretion of juice also results from influences arising from the intestine after the food has passed through the pylorus. This is the *intestinal phase of gastric secretion*. The cephalic and gastric phases account for about 45 per cent each of the total secretion of gastric acid, and the intestinal phase for the remaining 10 per cent.

THE PSYCHIC OR CEPHALIC PHASE

The psychic secretion of gastric juice was originally demonstrated by Pavlov, in dogs. To study in detail this phase of gastric secretion and other problems in gastric physiology he devised the following operation. A longitudinal incision was made in the body of the stomach in the region of the greater curvature (fig 39 1, AB). This included both anterior and posterior walls of the organ, and starting near the pyloric region ended about the middle of the fundus. The concave flap formed in this way, and which remained attached to the fundic region by its base, was then turned down and its cut edges, as well as those of the main stomach, sutured together. Thus was fashioned a small tube lined by normal gastric mucosa with an open free end (fig 39 1S). Complete isolation of the cavity of the tubular pouch from the main cavity of the stomach was accomplished by reflecting a flap of mucosa from the gastric walls at the junction of the pouch with the fundus and fixing it in position to form a partition between the two cavities. The free, open end of the pouch, or

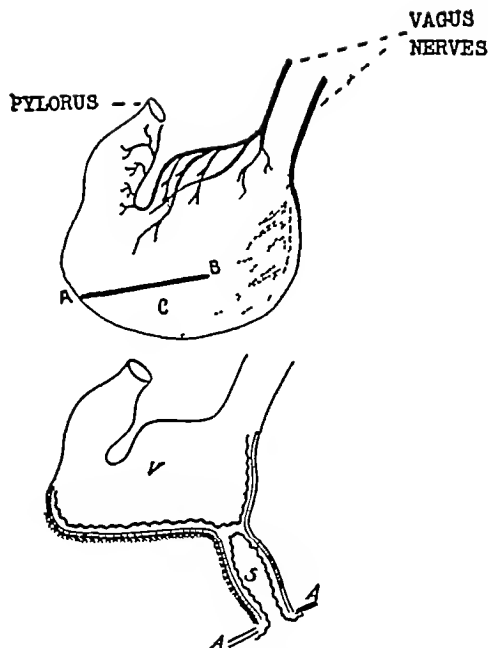


FIG 39.1 The Pavlov pouch. Upper drawing shows line of incision to form a flap C, cardiac part of stomach. Lower drawing shows the completed operation S, pouch, the mucous membrane has been reflected to form a partition between the main cavity of the stomach and the miniature stomach A, A, abdominal wall (after Pavlov)

miniature stomach, as it is usually called, was then sutured into the abdominal wound (fig 39.1, AA). Through this opening pure gastric juice uncontaminated by food could be collected while digestion was proceeding normally in the remainder of the stomach. The main stomach was also fastened to the abdominal wall and a fistulous communication established between it and the exterior. The advantage claimed for this operation over previous ones contrived for a similar purpose (e.g., the Heidenhain pouch) is that, on account of the longitudinal direction of the incision, a minimum amount of damage is inflicted upon the nerve supply of the pouch.¹ Furthermore, a comparatively large pouch—about one-tenth the size of the whole stomach—is procured by this means. The miniature stomach is assumed to portray faithfully

the behavior of the main stomach under various experimental conditions. This assumption has been justified by control experiments in which the secretion of the miniature and of the main stomach were compared.

Sham feeding

In the study of the *psychic* phase of gastric secretion it is, of course, essential that no food shall be allowed to enter the stomach. In order to secure this the esophagus was divided in its upper third, the lower end of the upper section was then brought out through the skin wound and fixed in position by sutures (fig 39.2). A miniature stomach and a gastric fistula leading into the main portion of the stomach were also produced, in the manner already described, for the purpose of observing the secretory responses, as well as for feeding the animal when necessary. An animal prepared in this way could eat and enjoy its meals but the food after being swallowed simply issued from the esophageal opening in the neck. When "*sham or fictitious feeding*", as this procedure is called, is performed a profuse secretion of gastric juice follows after a brief latent period. In some experiments in which sham feeding was maintained for 5 or 6 hours, as much as 700 cc. of pure gastric juice were secreted into the stomach. A proportionately smaller amount was secreted into the artificial pouch. Even after a single 5-minute period of fictitious feeding the secretion persists for from one to three hours, and produces 200 or 300 cc. of juice. The secretion is rich in pepsin. Division of the vagi completely abolishes the response. It is, therefore, undoubtedly due to a reflex mediated through these nerves. The quantity of juice varies with different foods, but the only quality of the food which appears to influence the psychic secretion is its palatability, the chemical or physical properties of the food are unimportant in this regard. Meat for instance for which the animal has a keen relish produces a juice of the greatest abundance and richest in ferment. Bread or other materials not particularly appetizing cause less secretion, and the juice has less digestive power. A more pronounced reaction occurs if the animal is hungry. In marked contrast to the activity of the salivary glands, no secretion can be evoked by placing inedible substances in the mouth, or materials which are injurious or nauseous, such as acid, pepper, asoefetida, etc. Indeed these may cause inhibition of a secretion that is already in

¹ This claim is unwarranted. Other methods have been devised which are more successful in preserving the vagal innervation. One of the best of these is the method of Thomas, in which a small incision is made into the stomach near the fundus through which strong toothed forceps are inserted and, seizing the interior of the gastric wall, draw part of the stomach through the incision to form a pouch (Thomas, J. E. Proc Soc Exp Biol and Med, 1942, 50, 58).

progress All these facts point to the involvement of the higher nervous centers in the reflex mechanism The pleasure of eating, the agreeable stimulation of the organs of taste and the gratification of appetite are essential conditions

It is not even necessary that the food should enter the mouth in order to elicit the gastric response Provided the food is sufficiently appetizing simply sight or smell of it will cause secretion In other words a *conditioned reflex* can become established for gastric as well as for salivary secretion (ch 69)

The importance of the psychic secretion may be judged from the fact that in its absence gastric digestion may be seriously hampered When, for example, a piece of meat or bread is placed in the stomach through a fistula while an animal is sleeping or, by diversion of its attention, is unaware of the introduction of the food, much less is digested in a given time than when the animal has been shown the food before introducing it into the stomach or has been given it by "sham feeding"

Babkin has shown that a time element may be prominent in the conditioned response, he found that animals which were accustomed to being fed at a certain hour each day secreted large amounts of gastric juice at this time though no food was offered

The psychic secretion in man

In man, the cephalic phase causes the secretion of from 50 to 150 cc within 20 minutes (Ivy) Richet observed long ago that in a subject who had suffered esophageal stricture and into whose stomach an artificial opening (fistula) had been made for feeding purposes, the secretion of gastric juice occurred when food was taken into the mouth Within more recent years Carlson has carried out extended observations upon a subject who had had a similar operation performed on account of an obstructed esophagus resulting from the ingestion of a corrosive in childhood This subject therefore was already prepared, like Pavlov's dogs, for sham feeding, the only difference being that there was no opening in the neck, it being necessary to spit out the food after chewing it All the findings of Pavlov were, in the main, confirmed in the case of this subject The influence of appetite or the desire for the food was particularly well brought out The subject especially enjoyed the dessert, and the curve of gastric secretion showed the most pronounced rise when sweets or fruit,

such as oranges, were chewed The secretion, however, did not last for as long as in Pavlov's experiments, but commenced to decline as soon as the stimulation of the taste buds had ceased (fig 39 3) Nor did the mere sight or smell of food (conditioned stimulus) evoke a response except when in one instance the subject was sent from the laboratory to select his meal from a near-by cafeteria In this instance there was a very definite response to the sight and smell of the food Several other observers have reported the existence of definite conditioned reflexes for gastric secretion in man

The psychic effect upon gastric secretion has been demonstrated also by Bennett and Venables upon subjects during hypnosis A suggestion made

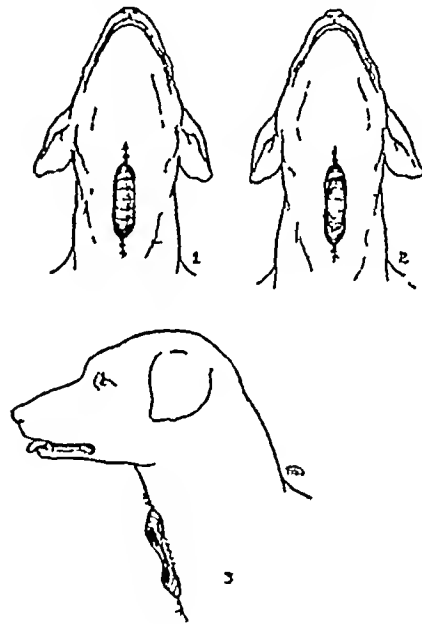


FIG 39 2 A two-stage operation for making an esophageal fistula 1 First stage, showing esophagus exteriorized 2 Second stage, performed four or five days later, showing excision of elliptical segment of anterior esophageal wall 3 Completed operation, lateral view (After Dragstedt and associates)

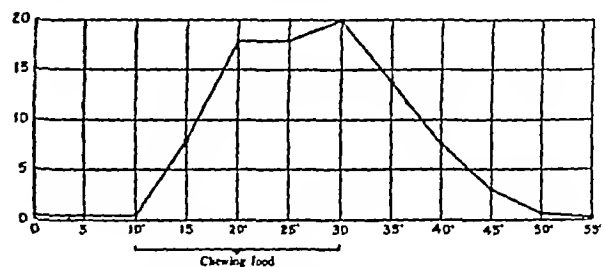


FIG 39 3 Typical curve of secretion of gastric juice collected at 5-minute intervals during the mastication of palatable food for 20 minutes The rise in secretion during the last 5 minutes of mastication is due to chewing the dessert (fruit) for which the person had an especial fondness (After Carlson)

to the subject that he was eating a savory food called for a secretion of juice. The mere suggestion of nauseating substances inhibited secretion (see below). In experiments upon medical students Hawk and his associates showed that a meal which had a disgusting appearance and a foul smell (indol was scattered over a dirty tablecloth) retarded digestion.

The application of these experimental results to dietetics is obvious. Foods agreeably flavored and attractive in appearance, impressions received from a meal prepared in a pleasing way, and, probably also, sensations aroused by the surroundings but not directly concerned with the food itself, have all an effect upon gastric secretion. The impulse which guides the gourmet is sounder physiologically than that which impels the glutton. The question of calories and the relative digestibilities of the various foodstuffs should not be allowed to obscure the psychic element entirely in these matters, for the "delights of the table" have true digestive value. These facts are expressed in the words of Pavlov "appetite spells gastric juice" or in the hospitable words of Macbeth "Now good digestion wait on appetite and health on both." Good gastronomical custom seems guided by this truth, for it has decreed that the meal shall begin and end with the more strongly flavored and appetizing dishes.

The effect upon gastric function of psychic states other than those associated with appetite

The gastric sensations so often associated with emotional states, excitement, fear, anxiety etc., are familiar to us all. That such subjective reactions are accompanied by changes in gastric function, e.g., secretion, vascularity and motility, which can be measured by suitable means, will occasion no surprise.

Hawk and his associates, for example, observed that anxiety or worry prolonged digestion. Wolf and Wolff have studied the secretion and motility of the stomach as well as the changes in vascularity of the mucosa in a patient with a gastric fistula during several types of emotional upset. Fear, mental depression or sadness, was associated with pallor of the mucous membrane and reduced secretion, whereas anxiety, with resentment, anger or aggressiveness, caused hyperemia and hypersecretion. Nervous tension, with anxiety and apprehension over a period of two or three weeks, corresponded during this time with hypernormal

vascularity and secretion. Generally, flushing or pallor of the face of this subject indicated a corresponding state of the gastric mucosa. It is to be expected, however, that since the sensitivity of facial vasomotor responses varies considerably in different persons, the vascular responses of the gastric mucosa would also show individual variations.

THE GASTRIC PHASE THE SECRETION OF
GASTRIC JUICE CAUSED BY FOOD
IN THE STOMACH

When food is present in the stomach the secretion of gastric juice continues for a much longer time than can be accounted for by the psychic secretion. The factors concerned in this secretory phase must now be considered. In the first place, "What is the nature of the adequate stimulus? Is it mechanical—the mere contact of the food with the gastric mucosa? Or is it chemical?"

(a) MECHANICAL STIMULI Pavlov demonstrated, apparently quite conclusively at the time, that mechanical stimulation of the gastric mucosa was entirely without effect upon the secretion of gastric juice. A glass rod having its tip covered with blue litmus paper remained unaltered in color when rubbed vigorously over the mucosa of the stomach. It was wetted by alkaline mucus but there was evidently no secretion of acid. Streams of sand when blown with considerable force against the mucosa so as to cause a more diffuse stimulation were also ineffective. More recently, however, Ivy and Farrell have shown that some degree of glandular excitation is caused by mechanical stimulation—particularly by the application of a distending force. There is an unusually long latent period and this is probably the reason that the effect was not discovered previously. The amount of juice secreted is, however, not great and, therefore, this type of stimulus cannot be wholly or even largely responsible for the flow which food induces by its presence within the stomach. It has also been shown that bone fragments, but not bone which has been powdered, when introduced into a dog's stomach stimulates secretion.

(b) CHEMICAL STIMULI It was observed by Pavlov that certain articles of food such as bread and meat, which could not differ to any significant degree from one another in their physical effects upon the gastric mucosa, produced, nevertheless, greatly different secretory responses. Bread introduced through a fistula into a stomach that had

been thoroughly dried of all residual juice, and the vagus nerves to which had been cut (in order to eliminate the psychic secretion), caused no excitation of the glands. Meat, on the other hand, under the same conditions, evoked an abundant flow of juice. The contrary results in the two instances must have been due, it was argued, to chemical differences between the respective foods. The adequate stimulus was evidently chemical in nature and dependent upon substances contained in the meat. The various constituents of meat—its salts, water, protein and extractives—were then investigated each in its turn for their secretagogue effect. The material was introduced into the main stomach, while the miniature stomach was observed for any signs of secretion.

With the exception of the *extractives* none of these components of meat had any pronounced effect. The *ash of meat* was without any effect whatever. *Water* caused a slight secretion (Rehfuß has shown that water causes a much greater excitation in human stomach than Pavlov reported for dogs). *Protein* itself, as in egg white, when the effect of its contained water was taken into account, appeared to be quite inert. But meat broths, meat juices and commercial extracts, such as Liebig's, all of which are rich in the watery extractives of meat, elicited a copious secretion.

Not only meat extractives but the *products of protein digestion*—*proteoses* and *peptones*—were found powerfully stimulating to secretion.

Having determined the specific stimulants for the gastric phase of gastric secretion, namely meat extractives and the products of peptic hydrolysis, the next question to be answered is "Through what mechanism do these substances act?" There are several possibilities. They might conceivably act (1) by direct stimulation of the gland cells, (2) through the stimulation of the afferent nerves in the gastric mucosa, thus bringing about secretion through a central reflex, (3) through their absorption into the blood stream, and their carriage therein to the glands, (4) through the local nerve plexus in the gastric wall, or, finally (5) by the liberation of a hormone from some part of the gastric mucosa which would then serve as the ultimate excitant.

The first two of these possibilities are ruled out as primary factors and (3) and (4) are improbable because of the following observations. Excitation of the glands of the miniature stomach occurs when the stimulating substances are placed in the main stomach. Injection of the materials intra-

venously causes a relatively small secretory response. The secretion occurs after the stomach has been completely isolated from the central nervous system. The fifth possibility remains to be considered.

*A consideration of a hormonal mechanism as
responsible for the gastric phase
The gastrin theory*

In 1902 Bayliss and Starling discovered a hormone (secretin, p. 530) in the upper part of the small intestine which had a very powerful effect upon the secretion of pancreatic juice. The discovery of this substance led Edkins a few years later to search for one of a similar nature which might be concerned in the control of gastric secretion. He found that when the pyloric mucosa was ground up and extracted with some of the stimulating substances mentioned above, notably peptones, or with glucose or hydrochloric acid, the extract had a powerful secretory effect when injected intravenously. A simple watery extract was ineffective. Injection of the substances themselves directly into the blood stream was also incapable of causing secretion. Edkins gave the name *gastrin* to the undetermined principle in the active extracts. The conclusion drawn from these results was that gastrin, formed during normal digestion through the action upon the pyloric mucosa of substances derived from the food, was absorbed into the blood stream, and upon reaching the fundic glands excited them to activity.

As a result of later work (Koch, Luckhardt and Keeton) it was found that pyloric extracts were not specific but in common with other tissue extracts contained a secretagogue principle. This observation apparently deprived Edkin's results of any physiological significance in so far as gastric digestion was concerned, for Popielski in 1920 had demonstrated that histamine, which may be isolated from various tissues, was a powerful stimulant to gastric secretion. *It was therefore suggested that the active principle in pyloric extracts (gastrin) was simply histamine.* This view has been amply substantiated. Histamine is a constituent of pyloric extracts, and Sacks, Ivy and collaborators found that if the pyloric tissue were incubated with histaminase (the histamine-inactivating enzyme) before extraction, the preparation had no secretagogue action. This observation has been confirmed by Gavin, McHenry and Wilson who have also shown that in the dog, the tissue of the body of the stomach is richer in histamine than that of

the pylorus and that 80 per cent of the total histamine content of the stomach is to be found in the former situation. Extracts of the fundus (body) have also a greater secretagogue action than have pyloric extracts.²

On the other hand, it has been shown definitely by Ivy and Farrell that an excitatory substance of some sort is actually carried in the blood stream, during digestion, to the glands of the fundus. These observers in a three-stage operation upon a dog, isolated completely a portion of the fundus and transplanted it into the subcutaneous tissue of the abdomen. After the operation, no nervous or direct vascular communications between the two sections of the stomach remained. When the animal was fasting, the transplanted pouch secreted a little mucoid fluid containing a small amount of combined acid, but no free acid. After the animal had been fed and digestion had proceeded in the main stomach for from 2 to 5 hours, an increase in the volume and in the combined acid of the juice secreted by the transplanted pouch occurred, in some instances free acid appeared.

The results of later experiments by Gregory and Ivy on dogs point to the blood-borne substance being a hormone. They created two completely denervated stomach pouches. One pouch, consisting of a small part of the fundic region, was transplanted (transplanted pouch) to the subcutaneous tissue of the mammary gland. The rest of the stomach (main pouch) was separated from the esophagus and duodenum which were then joined by anastomosis. Perfusion of the cavity of the main stomach with liver extract, which is a potent source of secretagogues, caused the secretion of free acid from both pouches. Mechanical stimulation (distension) of one pouch caused secretion from that pouch but not from the other. It is evident that, since the main pouch and the transplant are connected only through the general circulation, the action of the secretagogues is followed by the passage into the blood of a substance which stimulates the gastric glands, but that mechanical stimulation is incapable of initiating such a mech-

² The term *secretagogue* may be applied to any substance which excites secretion no matter by what mechanism the secretion is ultimately brought about. The term *humor* will be used to indicate a secretory excitant which is absorbed from the cavity of the stomach or intestine and carried in the blood stream to the gastric glands. The use of the word *hormone* will be confined here to any secretory substance formed in the gastric or intestinal tissue and conveyed to the glands in the blood stream.

anism. When the animal was given a meal of bread, meat and milk (which passed directly into the intestine) both pouches secreted free acid in fairly large amounts for several hours.

These results still left unsettled the question of the nature—humoral or hormonal—of the blood-borne substance, that is, whether the secretagogues themselves were absorbed into the blood stream or whether they acted upon the gastric mucosa to produce a new stimulating substance or to cause a *preformed* substance to be liberated into the circulation. Anesthetizing the mucosa of the main pouch with procaine prevented the action of the secretagogues in this part of the stomach as well as in the transplanted pouch. Yet after anesthetization of the transplant but not of the main pouch, secretagogues exerted their usual effect when introduced into the latter. This proves that the anesthetic does not paralyze the *gland cells*. Nor does procaine prevent the action of the blood-borne substance formed during normal digestion, once it has entered the blood stream, because secretion from the anesthetized main pouch occurs after the animal has been fed and food has entered the intestine. These experiments pointed definitely to a hormonal mechanism, although the possibility remained that the secretagogues themselves were humoral agents, the absorption of which had been depressed by the anesthetic. In experiments with other secretagogues such as histamine and ethyl alcohol, no evidence of such an effect of the anesthetic was secured.

In the still more recent experiments of Grossman, Robertson and Ivy who transplanted a fundic pouch larger than that used in the researches of Gregory and Ivy, a definite secretory response was obtained when the pyloric pouch was distended. The amount of secreted acid was greater during distension than in the control (predistension) period by 13 mg of HCl per hour, on the average (fig 394). Since in these experiments there is no question of the absorption of a secretagogue, the effect of mechanical stimulation is taken as an indication of a hormonal secretory mechanism brought into action by a mechanical stimulus. When this experiment and earlier ones are reviewed together there appears to be little doubt that chemical materials—secretagogues—act in a similar fashion.

The pyloric region has been thought by several to be essential for the production of the gastric hormone, but, in other experiments of Gregory and Ivy, pouches from which the pylorus had been

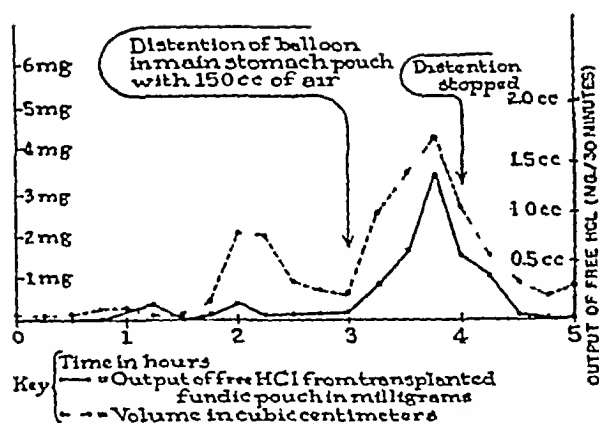


FIG. 39.4 Record of the secretion of acid by a transplanted fundic pouch of a dog when a pouch made of the remainder of the stomach was distended (After Grossman and associates)

removed gave essentially the same response when secretagogues were introduced into the remaining part of the stomach. The pyloric region, therefore, appears to have no special function in this regard.

THE NATURE OF THE GASTRIC HORMONE The question now arises whether or not the gastric hormone, i.e., the substance liberated into the blood stream when food enters the stomach, is identical with the active principle in pyloric extracts (gastrin).³ In other words, is histamine the gastric hormone? This question cannot at present be answered definitely. Though an increase in the histamine concentration of the blood has never been shown to occur during digestion, this negative result cannot be taken to disprove that histamine acts as a secretory hormone, for the amount required to stimulate the glands may be too small to be detected. That the secretory activity of pyloric extracts is destroyed by histaminase (p. 296), has been looked upon as important evidence that the principle contained in such extracts is simply histamine, but it does not disprove that a gastric hormone other than histamine and, *unobtainable by extraction*, exists. Furthermore, histaminase is not truly specific for histamine, therefore the enzyme might have destroyed a secretory stimulant other than histamine. Again, most histamine antagonists do not inhibit gastric secretion. The observations that an injection of atropine or the application of an anesthetic to the mucosa prevents the effect which follows the placing of food in the stomach but not the effect resulting from the injection of a pyloric extract or

of histamine, would seem to show that a principle other than histamine is produced in response to food. It is conceivable, however, that these drugs merely interfere with the *formation* of the hormone (whether it be histamine or some other substance) and yet have no effect upon the action of any hormone once it had been formed and entered the circulation.

The fact that the juice resulting from histamine injections differs from normal gastric juice suggests, again, that the gastric hormone is not simply histamine. According to Babkin and his associates the gastric juice formed by the stomach of the dog as a result of the injection of histamine or of purified gastrin, though possessing a high acidity, has a low concentration of pepsin. Babkin also points out that though histamine is very readily extracted by acids, hydrochloric acid placed in contact with the pyloric mucosa does not stimulate gastric secretion, a 0.5 per cent solution is actually inhibitory (Alley). Moreover, histological examination of the gland after histamine injections shows an effect confined to the acid-producing cells, in contrast to the appearance after vagal stimulation, the peptic cells were entirely unaffected (Bowie and Vineberg). It is stated, in opposition to these findings, that pepsin is present in fairly high concentration in human gastric juice produced by histamine injections (Bloomfield and Pollard), but in observations upon the human subject the possibility of the enzyme secreted prior to the injection having been simply washed out of the tubules by the secretion of the parietal (acid) cells is difficult to exclude. Indeed, evidence that the high concentration of pepsin was due to washing out of preformed ferment has been obtained by Toby who found that when histamine was administered repeatedly to patients the peptic activity of the juice declined progressively with each successive injection. Also, when an injection of neutral red is given it appears in the gastric juice. If after the concentration of dye in the juice has declined, gastric secretion is stimulated by histamine, the concentration rises sharply again, as also does the peptic activity. Washing of the dye from the tubules by the increased flow of secretion readily explains this result.

The question of the chemical (humoral or hormonal) control of gastric secretion is far from settled. All that has been established, so far, is that contact of secretagogues (e.g. meat extracts) with the gastric mucosa, especially of the pyloric

³ The term "gastrin" is used throughout this text to indicate exclusively the active substance (histamine) obtained by Edkins and subsequent workers in extracts of the pyloric mucosa.

region, or mechanical stimulation, such as distension, causes a principle to be absorbed into the circulation and carried to the gastric glands. The weight of evidence points to a hormone, which is not histamine, as being the active agent⁴.

In abnormal gastric states, e.g., inflammatory conditions, trauma and irritative agents may cause the liberation of histamine as they do in the skin and other tissues. The histamine thus released will act as an excitant of the glands. Such a release of histamine would explain why in persons with duodenal ulcer the interdigestive secretion of acid is not abolished by a dose of atropine which is effective in a normal subject.

It has been suggested that histamine may play a rôle in gastric secretion somewhat different from that just discussed, namely, by serving as a chemical mediator of the vagal control of the parietal cells, i.e., as a "local hormone" or humor liberated by the vagal nerve endings. The facts that histamine is present in the gastric juice, that its concentration in the juice is increased by vagal stimulation and that its content in the body of the stomach is much higher than that of the pyloric region accords with this possibility.

THE INTESTINAL PHASE OF GASTRIC SECRETION

The products of gastric digestion upon entering the duodenum act as chemical excitants to gastric secretion. If a meal of bread and milk is fed to an animal, as prepared in Gregory and Ivy's experiments, a secretion from both the main stomach pouch and the transplant commences within about 2 hours and continues for from 3 to 9 hours. It appears that the mechanism is humoral in nature, and similar to that which stimulates secretion during the gastric phase of digestion, split digestive products acting as secretagogues. Mechanical stimulation of the intestine, as by distension, is ineffective, and, as mentioned on page 510, the application of procaine to the gastric mucosa does not abolish the response. Various substances placed directly in the duodenum, e.g. water, meat extracts, albumoses and peptones, magnesium sul-

phate, saponin, soaps, etc., have been shown to excite the glands of an isolated stomach pouch. The importance of food in the duodenum as a stimulus to gastric secretion was shown by Crider and Thomas who found that if the duodenal contents were drained away through a fistula, the response of the gastric glands to a meal was only a third of the normal value.

INHIBITION OF THE GASTRIC GLANDS BY FAT
The inhibitory effect of fat upon gastric secretion, though known for years, was thought to be due to a nervous reflex. The effect is exerted only after the fat has passed through the pylorus. A secondary stimulating action of fat upon gastric secretion occurs, however, after it has undergone digestion, due to the production of soaps.

The *motility* of the stomach is also depressed by fat (p. 567), hunger contractions as well as peristalsis during digestion are inhibited. The manner in which fat exerts its inhibitory effects upon the gastric functions has been clarified by Ivy and by Lim and their associates. A hormone (chalone, ch. 57) is responsible. Fat placed in the duodenum was found to inhibit the motility and secretion of a denervated pouch of the entire stomach. Both acid secretion and peptic secretion were suppressed, the latter often to a greater degree than the former. The inhibitory effect is not due to absorbed fat, for it is not abolished by establishing a thoracic duct fistula and draining the chyle to the exterior, and injection of fat into the circulation or of any product of its digestion is ineffective. The discharge of bile into the duodenum is stimulating rather than inhibitory to gastric secretion. Cholecystokinin and secretin are also without any inhibitory action. An active principle was extracted by Ivy and his associates and named *enterogastrone*. This material is free from vasodilator (histamine) activity, and from both cholecystokinin and secretin.

Enterogastrone is capable in maximal dosage of causing complete suppression of secretion for from 1 to 5 hours and of gastric motility for 30 minutes. The latter effect is not evident upon a *denervated* isolated gastric pouch. The secretion of pepsin by a denervated pouch is also much less pronounced than that from one with nerves intact. Inhibition of *acid* secretion, on the contrary, is unaffected by denervation. As the inhibitory principle is more highly purified and concentrated, the inhibition of gastric motility becomes more pronounced, and the depressing effect upon secretion

⁴ Komarov states that he has prepared a pyloric extract free from histamine which stimulates the glands upon intravenous injection. Uvnas has more recently reported similarly. The secretion evoked by this extract resembled, however, that caused by histamine rather than normal gastric juice secreted during digestion. It was found to be very low in peptic power and was not affected by atropine. Freedman and King obtained an extract containing traces of histamine which stimulated gastric secretion. It is possible that the extracts obtained by these authors contained a split product of protein which caused the liberation of histamine from body tissues.

less so, an indication that the two effects are due to separate substances and not, as was originally thought, to a single one⁶

A substance known as *urogastrone* and having effects similar to enterogastrone upon gastric secretion and motility has been extracted from normal urine by Gray and his associates. Urogastrone was thought at first to be most probably excreted enterogastrone but certain differences have since been demonstrated which indicate that they are separate and distinct substances. Pepsin inactivates the motility-depressing action of enterogastrone but not of urogastrone. Also, urogastrone has a less prolonged depressing effect upon gastric motility than has enterogastrone, and excision of the stomach and duodenum does not abolish urogastrone from the urine of dogs, even removal of the entire small bowel though it reduces the excretion of urogastrone does not cause it to disappear from the urine.

THE ADAPTATION OF THE QUANTITY AND QUALITY OF THE GASTRIC SECRETION TO THE TYPE OF FOOD. It was shown by Pavlov and Khizhin that the juices secreted for the three foodstuffs, bread, meat and milk, respectively, differ characteristically from one another, both in quantity and digestive power. Indeed the juice secreted for each type of food is, according to Pavlov so specific and suited to the digestion of the particular foodstuff which calls it forth, that it is possible to predict the character of the juice which will be secreted when a given type of food is fed. He therefore spoke of "meat", "bread" and "milk" juice, respectively. The juice secreted for *meat*, for instance, was found to be greatest in quantity, but intermediate in digestive power between those formed for bread and for milk. *Bread juice* possessed the highest peptic power (about double that of meat juice) but was intermediate in amount. It was suggested that the high ferment value was an adaptation furthering the digestion of the more resistant vegetable proteins. *Milk juice* was both scanty in amount, and poor in peptic power during the first

⁶ The differences between the effects of fat in the duodenum and of the principle obtained by extraction lead one to believe that this so called *exogenous* enterogastrone is not the true hormone (*endogenous* enterogastrone) which mediates the inhibitory effect of fat upon gastric function. For example, though fat in the intestine affects the secretion of acid less than that of pepsin, the reverse is true for a preparation of enterogastrone, the latter sometimes appears to stimulate peptic secretion. Furthermore, whereas fat in the intestine is very effective in man in inhibiting gastric secretion, very large doses of exogenous enterogastrone are required to produce an effect. Again, enterogastrone does not, as already mentioned, inhibit motility of a completely denervated stomach pouch.

TABLE 36*

	SECRETION TOTAL	ACID CON- CENTRATION	PEPSIN CON- CENTRATION
Meat	High	High	Medium
Bread	Medium	Low	High
Milk	Low	Medium	Low

* Modified from Carlson

four hours of secretion. The low values of the juice secreted for milk were ascribed to the inhibitory effect of the milk fat. The acidity is highest in meat and lowest in bread juice. In the case of the human stomach meats are recognized to be, as compared with other foods, the most powerful stimulants of acid production (See table 36.)

The acidity of the gastric juice as collected from the stomach in many instances shows a tendency to vary with the rate of secretion by the fundic glands, when their rate of secretion is slow the juice will be diluted to a greater extent by a non-acid fluid than when the rate is rapid. It was Pavlov's view that variations in acidity are due entirely to changes in the rate of *total* secretion into the stomach and that *the acidity of the juice as secreted by the parietal cells is constant* (p. 498). The reverse relation holds between rate of secretion and pepsin concentration, the latter becoming reduced as the secretion rate increases. These relationships are dependent upon the fact that the acid secreted by the parietal cells makes up the bulk of the fluid of the juice. Juice of large volume is therefore highly acid, that of small volume contains a relatively high concentration of pepsin and alkaline fluid. Such relationships hold true, however, only when the same type of secretory stimulus is employed but its intensity varied. When, however, the type of stimulus is altered it appears that the acidity or peptic activity of the juice may vary independently of the rate of secretion. That is, a slow or rapid secretion rate may be associated with either a low or high acidity, and similarly with the pepsin concentration. Furthermore, the proportions of the three main constituents of the juice, acid, pepsin or mucin, can vary independently of one another.

THE EFFECTS OF VARIOUS CHEMICALS AND DRUGS UPON GASTRIC SECRETION

Alkalis in general have been held to exert a depressing effect upon the secretion of gastric juice. Sodium bicarbonate, for instance, a favorite ingredient of digestive mixtures was investigated by Pavlov in dogs and found to be definitely inhibitory. This observation has been confirmed by Farrell, though others find that the inhibitory effect does not occur unless the dose is excessive,

and that in small repeated doses it augments the secretion (Boyd) This seems to be true also of other alkalis, large doses depress but small doses, especially if repeated, may augment the flow of juice Nor does the inhibitory effect of large doses persist after their discontinuance, on the contrary, hypersecretion not uncommonly follows ("acid rebound") The value of alkalis in gastric disorders depends chiefly, however, upon their antacid properties, i.e., upon their ability to neutralize or buffer the acidity of the gastric contents rather than to depress secretion

The various preparations of *bitters* are without any appreciable effect upon secretion unless they contain alcohol *Acids* depress gastric secretion which is completely inhibited by the introduction into the stomach of a 1 per cent solution of hydrochloric acid (p 517) Acid also exerts an inhibitory effect from the intestine upon gastric secretion When introduced into the dog's intestine gastric secretion from a Pavlov pouch is inhibited *Condiments* have little direct effect, but act indirectly in adding flavor to the food, stimulating the taste buds and thus encouraging the psychic secretion *Histamine* is one of the most powerful stimulants of gastric secretion (p 518) Histamine liberated within the body e.g. in dermographism (p 315) or even by the immersion of the hand in cold water at a temperature of 10°C causes a detectable secretory response within 15 minutes Gastric secretion is not inhibited by such antihistamine drugs as *neonantergen* and *benadryl*, and most other agents of the same class, of the few which are antisecretory the dose required must be inordinately large *Caffeine* and *alcohol* are strong secretory stimulants The latter has a pronounced secretagogue action, causing the secretion of a juice of high acidity and rich in mucin It is possible that alcohol exerts its secretory action through the liberation of histamine, for it has been shown that the histamine output of the perfused lung of the guinea pig is increased by the addition of alcohol (2-6 per cent) to the perfusion fluid *Liver extract*, *meat*, and *vegetable extracts* generally are powerful excitants of the gastric glands, an action which they owe to the presence of secretagogues *Insulin* (through its hypoglycemic effect on the vagus center),⁶ *acetylcholine* (usually), *mecholyl*

⁶ Insulin fails to stimulate gastric secretion, or its effect is greatly reduced, after vagal denervation This fact is made use of as a test for the success of vagotomy in the treatment of duodenal ulcer The hypothalamus apparently contains chemoreceptors (glucoreceptors)

(*acetyl-β-methylcholine chloride*), *pilocarpine* and *nicotine* are secretory stimulants, while *belladonna* or its alkaloids *atropine* and *hyoscyne*, and *hyoscyamus* or its alkaloid *hyoscyamine* are secretory depressants Atropine or vagus section depresses or abolishes the cephalic secretion, the drug reduces somewhat, but does not suppress, secretion due to histamine, alcohol, or caffeine *Smoking* depresses secretion, while *morphine*, after a brief period of inhibition, stimulates secretion *Adrenergic blocking* agents do not affect secretion

THE EFFECTS OF OPERATIVE PROCEDURES UPON GASTRIC SECRETION

In dogs, partial gastrectomy (excision of the pylorus and anastomosis of the gastric stump to the duodenum) is followed by a pronounced reduction in the *quantity* of juice secreted by the fundic glands but no change occurs in the *concentration* of acid The total acid secretion is reduced to $\frac{1}{2}$ or $\frac{1}{3}$ (see Wilhelmj and associates) Mucous secretion is increased, and combining with the acid further reduces the free acid concentration of the gastric contents Partial gastrectomy, accompanied by section of the gastric vagi, results in a profound reduction in acid secretion in response to a test meal for both cephalic and gastric phases are depressed, section of the vagi alone, abolishes the cephalic phase, histamine, however, still causes a well-marked secretion

ANALYSIS OF THE STOMACH CONTENTS—TOTAL ACIDITY—FREE ACIDITY

During digestion the hydrochloric acid in the stomach exists in two forms—a smaller part combined with the protein of the food, regurgitated duodenal secretions, saliva, mucus and the secretion of non parietal cells—*combined acid*, and a much larger part present in the free state—*free acid* The value for the combined acid is, however, of little physiological or clinical importance The sum of the combined acid and free acid, together with acid salts and any organic acids, lactic etc that may be present, is termed the *total acidity* Meat and other protein rich foods are more capable of combining with hydrochloric acid than are carbohydrates When the hydrochloric acid is absent or falls to a low level, organic acids, as a result of the fermentative processes, tend to be formed

Samples of gastric contents are obtained for analysis by means of a Rehfuess or a Ryle's tube, the former is made of narrow flexible tubing fitted at one end with a which respond to the blood glucose level through which, and the vagal fibers, the insulin effect upon gastric secretion is mediated

metal tip possessing large perforations. The subject is instructed to swallow the tube slowly. To its upper end a glass syringe is attached, and by means of this the stomach contents are aspirated. This is done the first thing in the morning before food or drink has been taken. The healthy resting stomach secretes continuously and the fluid removed at this time (*residual juice*) under normal conditions amounts on the average to 50 cc. The total acidity of such a sample normally averages 30 units and the free acidity 20 units (see below). After the removal of the residual juice the gastric glands are excited to secrete by the ingestion of a *standard test meal*. Ewald's test meal is frequently employed and consists of two ordinary slices of toast (35 grams) and 8 ounces (250 cc) of weak tea without sugar or cream. Others give a pint of thin oatmeal gruel (Boas meal). Dilute alcohol (50 cc of a 7 per cent solution), a solution of meat extract (e.g., Liebig's), or an injection of histamine is also employed for stimulating gastric secretion. When the test meal is used a specimen of gastric contents (10 to 15 cc) is removed 15 minutes after the meal has been swallowed, and during the succeeding 2½ hours at 15 minute intervals. Finally, the gastric contents are completely removed. Determinations of the free and the total acid are then made upon each sample and the corresponding acidity curves constructed (fig. 39.5, p. 516). The tube remains in position throughout the test. It was at one time the custom to make a single analysis 1 hour after the meal and to draw conclusions concerning gastric function from the results of this. But, since the acidity is continually changing, the construction of a curve from the results of analyses made at short intervals is the only logical procedure. This is known as the *fractional method of gastric analysis*, and was introduced by Rehfuess and his collaborators.

Chemical tests

QUALITATIVE (a) Free hydrochloric acid A drop of Toepfer's reagent which consists of a 0.5 per cent solution of dimethyl amino azo-benzene is added to a cubic centimeter or so of the gastric contents. This reagent gives a red color at a pH below 3.0 that is in the presence of free HCl. It is not quite specific for it may react to high concentrations of organic acids. A yellow color on adding the reagent indicates the absence of free HCl.

(b) Lactic acid When free HCl is absent from the gastric contents (anacidity, p. 518) fermentation is likely to occur with the production of organic acids, particularly lactic. The latter is detected by means of Kelling's test—a modification of Uffelmann's. Two drops of 10 per cent ferric chloride are added to a test-tube of distilled water. The solution is divided between two test-tubes in equal portions. A few drops of filtered gastric contents are added to one tube, a lemon yellow color appears if lactic acid is present. The solution in the other test-tube serves as a standard for comparison. The gastric contents may contain other

substances which give the test and, for this reason, it is preferable to take up the lactic acid with ether, evaporate to dryness and dissolve the residue in water. The test is then applied to the aqueous solution.

QUANTITATIVE (a) Free hydrochloric acid One cubic centimeter of strained gastric contents is diluted with 10 cc of distilled water. Two drops of Topfer's solution are added as an indicator. The sample is then titrated with 0.01 N NaOH until the solution turns a salmon pink color (pH 3.3).

(b) Total acidity After noting the reading of the burette and estimating the free acid, three drops of phenolphthalein solution are added to the specimen and the titration with sodium hydroxide continued until the solution turns a definite pink. More accurate results are obtained with phenol red, the end point being at pH 7.

The results of these determinations are expressed in "clinical units" i.e., the number of cubic centimeters of 0.1 N sodium hydroxide solution which are required to bring 100 cc of stomach contents to the end-point in each case. The titration figures are, therefore, multiplied by 10. The total quantity of alkali required gives the value for the total acid, the first titration gives the free acid. The *percentage* of free HCl is obtained by multiplying the number of 0.1 N cubic centimeters of alkali required for neutralization by the factor 0.00365 (see graph, fig. 39.5).

Estimation of peptic power

The earliest method of determining the peptic activity of a sample of gastric juice was introduced by Mett in 1889. Egg white is drawn into a glass tube of 1 mm bore and a few inches long. The albumin is coagulated by placing the tubes in water at a temperature of 85°C and leaving them until the water has cooled. The tubes are then broken into sections about an inch long, and immersed in the gastric juice diluted 1 to 15 with N/20 HCl and incubated at a temperature of 37°C for 24 hours. After this time the length of the column of digested albumin at each end of the tubes is measured in millimeters under the low power of the microscope and the average of a number of measurements taken. The peptic power of the sample is expressed in accordance with Schutz's law which states that the amount of proteolytic enzyme present is proportional to the square of the number of millimeters of digested albumin. Therefore, if the average length of the digested columns is 2.5 mm the peptic power of the undiluted juice will be $2.5^2 \times 16 = 100.0$. This is the average normal value.

Anson and Mirsky's method utilizes hemoglobin as the substrate. After the juice has acted upon the hemoglobin for 10 minutes, the undigested protein is precipitated with trichloroacetic acid and removed by

⁷ The figure obtained corresponds to the milli normal concentration (mN) or milliequivalents of acid per liter.

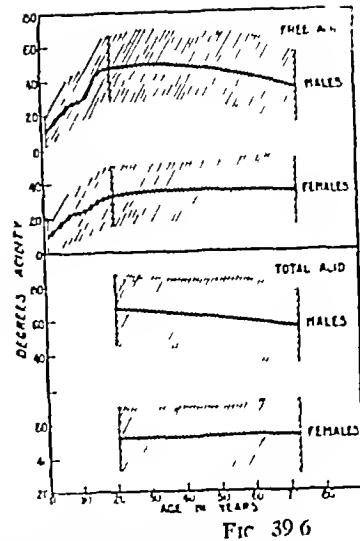
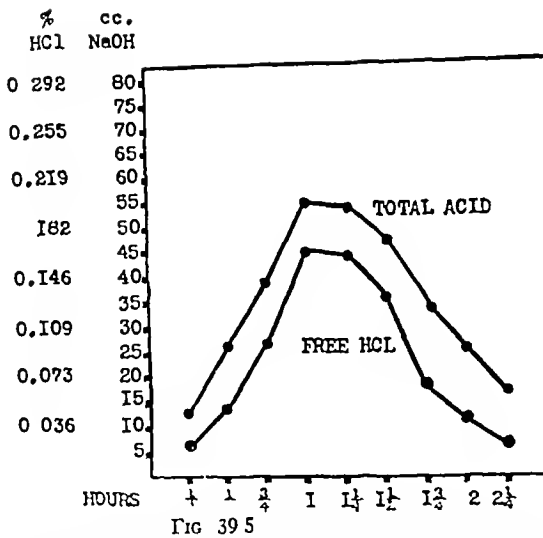


FIG 39 5 Normal curves of gastric acidity following a test meal (Redrawn from Maclean)
FIG 39 6 Normal standards The shaded areas represent the limits within which lay 80 per cent of the data for free and total acid at the different ages The heavy lines represent modes (After Vanzant, Alvarez and associates)

filtration The quantity of protein split products, which is a measure of peptic activity, is estimated colorimetrically after the addition of phenol reagent

Determination of gastric secretory function

It is becoming more and more evident that the methods of gastric analysis as carried out in the time honored fashion are inadequate as a means of gaining knowledge of the secretory activity of the stomach They measure the acidity of the gastric contents and, therefore, frequently give false information or equivocal results concerning secretion The acidity of the gastric contents removed after a test meal depends not only upon the secretion by the gastric glands but also upon the degree to which the secretions have been diluted by the fluids of the test meal still remaining in the stomach, as well as by saliva and fluids regurgitated from the duodenum It is not possible to know how much of the meal has left the stomach by the time the samples are withdrawn, unless there be added to it some soluble substance which is neither absorbed nor destroyed in the stomach and can be easily detected Phenol red may be used for this purpose The proportion of the gastric contents which has been secreted can then be determined and by dividing the value for the total acidity by the volume of the secretions, the acid concentration of the gastric juice itself (which contains a variable proportion of mucin) can be estimated Furthermore, since the acid concentration of the parietal cell secretion is constant, it is possible to determine the proportions of acid and non acid secretions in the juice

The test meal of bread or toast and tea usually employed is unappetizing and too weak a stimulus for all phases of gastric secretion nor can alcohol or histamine

rightly be considered a normal physiological secretagogue Liebig's test meal, consisting of a 2 per cent solution of Liebig's meat extract, contains potent and physiological secretagogues and is probably the nearest approach to the ideal test meal The histamine test (p 518) is valuable, however, as a means of distinguishing true from false acidity

The quantity of the juice secreted in the psychic or cephalic phase alone can be determined by having the subject chew for a time some appetizing food, such as an orange, and then reject it But such a procedure is rarely necessary for, since the continuous secretion of the fasting stomach is psychic in origin, the withdrawal of the gastric contents at intervals from the empty stomach is usually sufficient (see also neutralization test, p 520)

THE ISO SECRETORY OR NORMAL CURVES OF GASTRIC ACIDITY

If in a normal person the free and total acidities of the gastric contents be determined every 15 minutes for a period of from two to three hours after the ingestion of a test meal, the results plotted against time along the base line with HCl percentages or clinical units in the vertical axis, a curve is obtained as shown in figure 39 5 The curve for total acidity commences to rise a short time after the meal, and about 1 hour later reaches a maximum which varies from 35 to 70 in different persons The curve maintains its maximal height

for half an hour or less and then commences to decline reaching the resting level again in from $2\frac{1}{2}$ to 3 hours after the ingestion of the test meal. The curve of free acidity runs parallel to, but at a lower level than that for the total acidity, the values ranging in different normal persons between 20 and 40 (0.07 to 0.15 per cent). Values are much higher after foods, such as meats, which stimulate gastric secretion more powerfully, averaging from 80 to 120 for total and from 60 to 100 for free acidity.

The figures given above represent the range of the great bulk of normal persons, but the gastric acidities show very wide individual variations in health, being influenced markedly by age and sex. The average free and total acidities (after a test meal) in young healthy males are about 40 and 65, respectively (see figs 39.6 and 39.7). They are somewhat lower in females. They are also lower in children, the adult level being reached at about the age of 20 years. In men after 30 years and in women after 50 years a progressive decline in acidities (total and free) occurs and the incidence of anacidity and of subacidity (p. 518) increases sharply. A high normal value for gastric acidity is regarded by some as an index of physical fitness, the level tending to be low, it is said, in persons of sedentary habits and poor muscular development.

THE REGULATION OF THE ACIDITY OF THE GASTRIC CONTENTS

The normal acidity curves, as we have seen, reach their maxima in about an hour and then commence to decline. The fall in the acidity of the gastric contents has been attributed by various authors to the following factors:

- (1) Reduction in the concentration of acid in the parietal juice
- (2) Reduction in the total volume of juice as a result of the cessation of the psychic stimulation and gradual reduction in the secretagogue action of the food as gastric and intestinal phases subside
- (3) Evacuation of the stomach
- (4) Reduction in free acidity of the gastric contents by neutralization and dilution with saliva, and a non-parietal alkaline solution from fundic and pyloric glands, as well as by food derivatives and mucus from the surface epithelium
- (5) Neutralization through the action of urease
- (6) Neutralization of the juice by the regurgitation of alkaline duodenal fluids—especially pancreatic juice—into the stomach

With regard to the first factor, some authors

(Rosemann, and Maclean and Griffiths) have thought that the parietal cells could secrete chloride ions in combination either with Na or H ions, i.e., as NaCl or HCl, the proportions of acid and of neutral chloride in the juice being determined by the H ion concentration of the gastric contents, a rise in their acidity causing automatically inhibition of the secretion of H ions and a larger proportion of chlorine to be secreted as the neutral salt. However, the original view of Pavlov that the acidity of the gastric juice as secreted by the parietal cells remains constant has been confirmed by Hollander and is now widely accepted. It is apparently only the volume of the parietal secretion which varies. A rise in the acidity of the gastric contents automatically reduces the quantity of juice produced during the gastric phase of secretion. A high acidity in the duodenum also inhibits both the gastric and intestinal phases of secretion. The cephalic phase remains unaffected by acid either in the stomach or duodenum. The inhibitory effect of acid appears to be mediated through a nervous mechanism rather than by a blood-borne agent.

The passing of the cephalic phase and the weakening of stimuli in other phases of secretion as the stomach is evacuated, together with the neutralization and dilution by a non-parietal fluid (4 above) are probably the most important factors governing gastric acidity. The urea-urease-ammonia mechanism (p. 522) also is probably of major importance.

Regurgitation of fluids from the duodenum has been considered an important contributory element in lessening acidity. But the importance of this factor has been questioned by some investigators (Hollander, Shay) and affirmed by others (Wilhelmj, and Bolton and Goodhart). That duodenal regurgitation occurs is an undoubted fact as is

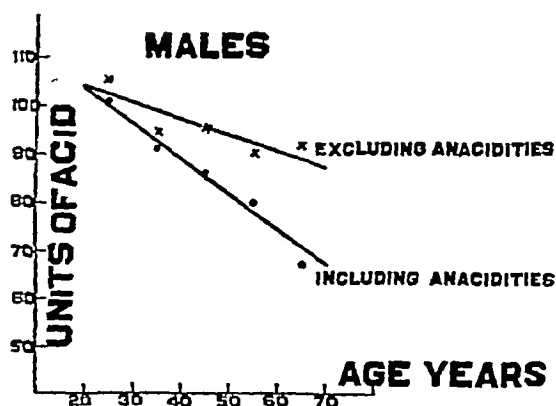


FIG. 39.7 Percentage decline of mean gastric acidity with advancing years (males) (After Pollard)

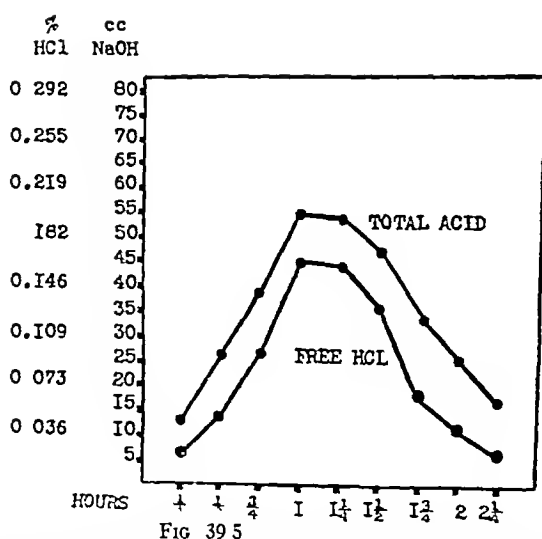


FIG 39 5

FIG 39 5 Normal curves of gastric acidity following a test meal (Redrawn from Maclean)

FIG 39 6 Normal standards. The shaded areas represent the limits within which lay 80 per cent of the data for free and total acid at the different ages. The heavy lines represent modes (After Vanzant, Alvarez and associates)

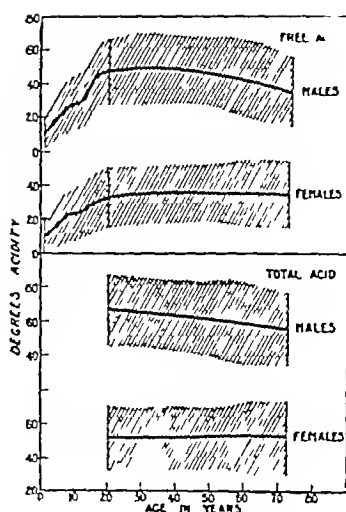


FIG 39 6

filtration. The quantity of protein split products, which is a measure of peptic activity, is estimated colorimetrically after the addition of phenol reagent.

Determination of gastric secretory function

It is becoming more and more evident that the methods of gastric analysis as carried out in the time honored fashion are inadequate as a means of gaining knowledge of the secretory activity of the stomach. They measure the acidity of the gastric contents and, therefore, frequently give false information or equivocal results concerning secretion. The acidity of the gastric contents removed after a test meal depends not only upon the secretion by the gastric glands but also upon the degree to which the secretions have been diluted by the fluids of the test meal still remaining in the stomach, as well as by saliva and fluids regurgitated from the duodenum. It is not possible to know how much of the meal has left the stomach by the time the samples are withdrawn, unless there be added to it some soluble substance which is neither absorbed nor destroyed in the stomach and can be easily detected. Phenol red may be used for this purpose. The proportion of the gastric contents which has been secreted can then be determined and by dividing the value for the total acidity by the volume of the secretions, the acid concentration of the gastric juice itself (which contains a variable proportion of mucin) can be estimated. Furthermore, since the acid concentration of the parietal cell secretion is constant, it is possible to determine the proportions of acid and non acid secretions in the juice.

The test meal of bread or toast and tea usually employed is unappetizing and too weak a stimulus for all phases of gastric secretion nor can alcohol or histamine

rightly be considered a normal physiological secretagogue. Liebig's test meal, consisting of a 2 per cent solution of Liebig's meat extract, contains potent and physiological secretagogues and is probably the nearest approach to the ideal test meal. The histamine test (p 518) is valuable, however, as a means of distinguishing true from false anacidity.

The quantity of the juice secreted in the psychic or cephalic phase alone can be determined by having the subject chew for a time some appetizing food, such as an orange, and then reject it. But such a procedure is rarely necessary for, since the continuous secretion of the fasting stomach is psychic in origin, the withdrawal of the gastric contents at intervals from the empty stomach is usually sufficient (see also neutralization test, p 520).

THE ISO-SECRETORY OR NORMAL CURVES OF GASTRIC ACIDITY

If in a normal person the free and total acidities of the gastric contents be determined every 15 minutes for a period of from two to three hours after the ingestion of a test meal, the results plotted against time along the base line with HCl percentages or clinical units in the vertical axis, a curve is obtained as shown in figure 39.5. The curve for total acidity commences to rise a short time after the meal, and about 1 hour later reaches a maximum which varies from 35 to 70 in different persons. The curve maintains its maximal height

hour or less and then commences to bring the resting level again in from 15 to 30 minutes after the ingestion of the test meal. Free acidity runs parallel to, but at a lower level than that for the total acidity, the range in different normal persons between 10 to 15 per cent. Values are much lower for foods, such as meats, which stimulate secretion more powerfully, averaging from 20 to 40 per cent total and from 60 to 100 for free

acidity. The values given above represent the range of acidity in a group of normal persons, but the gastric acidity shows very wide individual variations in acidity, being influenced markedly by age and sex. Free and total acidities (after a test meal) in healthy males are about 40 and 65, respectively (see figs 39.6 and 39.7). They are lower in females. They are also lower in the young, the adult level being reached at about 20 years. In men after 30 years and in women after 50 years a progressive decline in total and free) occurs and the incidence of hypochlorhydria and of subacidity (p. 518) increases. The high normal value for gastric acidity is regarded by some as an index of physical fitness, and, being found to be low, it is said, in persons of advanced age and poor muscular development.

FACTORS INFLUENCING THE ACIDITY OF THE GASTRIC CONTENTS

Normal acidity curves, as we have seen, show a rise to a maxima in about an hour and then a gradual decline. The fall in the acidity of the gastric contents has been attributed by various investigators to the following factors:

1. A fall in the concentration of acid in the gastric juice.

2. A fall in the total volume of juice as a result of a cessation of the psychic stimulation and a consequent reduction in the secretagogue action of the gastric and intestinal phases subsiding after stimulation of the stomach.

3. A fall in free acidity of the gastric contents due to neutralization and dilution with saliva, or with a parietal alkaline solution from fundic glands, as well as by food derivatives which are neutralized from the surface epithelium.

4. A fall in the concentration of the juice by the regurgitation of alkaline duodenal fluids—especially pancreatic—into the stomach.

As regards the first factor, some authors

(Rosemann, and Maclean and Griffiths) have thought that the parietal cells could secrete chloride ions in combination either with Na or H ions, i.e., as NaCl or HCl, the proportions of acid and of neutral chloride in the juice being determined by the H ion concentration of the gastric contents, a rise in their acidity causing automatically inhibition of the secretion of H ions and a larger proportion of chloride to be secreted as the neutral salt. However, the original view of Pavlov that the acidity of the gastric juice as secreted by the parietal cells remains constant has been confirmed by Hollander and is now widely accepted. It is apparently only the volume of the parietal secretion which varies. A rise in the acidity of the gastric contents automatically reduces the quantity of juice produced during the gastric phase of secretion. A high acidity in the duodenum also inhibits both the gastric and intestinal phases of secretion. The cephalic phase remains unaffected by acid either in the stomach or duodenum. The inhibitory effect of acid appears to be mediated through a nervous mechanism rather than by a blood-borne agent.

The passing of the cephalic phase and the weakening of stimuli in other phases of secretion as the stomach is evacuated, together with the neutralization and dilution by a non-parietal fluid (4 above) are probably the most important factors governing gastric acidity. The urea-urease-ammonia mechanism (p. 522) also is probably of major importance.

Regurgitation of fluids from the duodenum has been considered an important contributory element in lessening acidity. But the importance of this factor has been questioned by some investigators (Hollander, Shay) and affirmed by others (Wilhelm, and Bolton and Goodhart). That duodenal regurgitation occurs is an undoubted fact as is

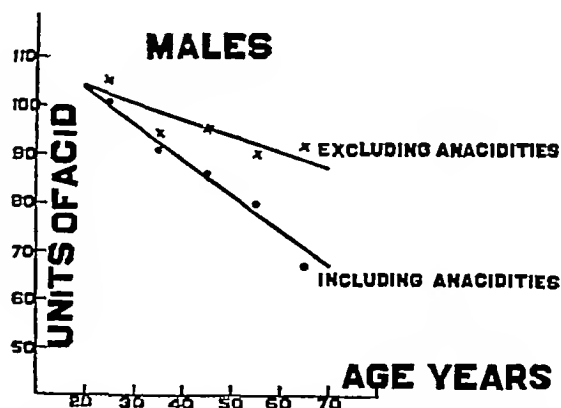


FIG. 39.7. Percentage decline of mean gastric acidity with advancing years (males). (After Pollard.)

evidenced by the appearance of bile and trypsin in the stomach, especially in the later stages of digestion Boldyreff in 1907 was the first to give duodenal regurgitation a prominent rôle in the regulation of gastric acidity This observer maintained that normally a high acidity of the chyme issuing from the pylorus stimulated the secretion of pancreatic juice and set up antiperistalsis in the duodenum which carried the alkaline fluid into the stomach The following observations, however, seem to show that duodenal regurgitation is a non-essential part of the mechanism controlling gastric acidity

(a) Baird, Campbell and Hern observed in human subjects that the usual fall in gastric acidity occurred though the reflux of fluid from the duodenum into the stomach was prevented (duodenal contents removed by duodenal tube)

(b) McCann prevented regurgitation in dogs by separating the duodenum from the stomach and jejunum, and draining it into the lower part of the ileum The stomach was then jointed to the jejunum The acidity curves of these animals did not differ from those of normal animals.

(c) Gastric acidity curves of isolated gastric pouches in dogs are essentially the same as those obtained from the intact human stomach

(d) Shay, Katz and Schloss found that alkaline fluids introduced into the stomach caused a greater reflux from the duodenum than did strong acids Water and weak acids were as effective as strong acids These authors also pointed out that normally the juices secreted into the duodenum have a pH of around 7.2 or lower, and consequently, are not sufficiently alkaline to exert, in the quantities which are regurgitated, an important effect upon gastric acidity Moreover, the

secretory curve of a fundic pouch of the dog isolated from the rest of the stomach is similar in character to that of the intact human stomach (fig 39 8)

THE HISTAMINE TEST

As a test of gastric secretion the parenteral administration of histamine although not a normal stimulus has certain very definite advantages over the ordinary test meal (a) Histamine evokes a maximum secretory response and is often able to evoke a response when the test meal fails to do so (see below) It is thus of value in distinguishing false from true acidity (b) The response is not affected by conditions such as appetite and other psychic factors which influence the response to the test meal (c) The test meal and salivary secretion, which latter cannot be measured, add to the volume of the contents, it is therefore impossible to determine accurately the quantity of juice secreted (d) Swallowed saliva and the test meal itself partly neutralize the acid (e) In the histamine test the glands respond promptly, maximal acidity being reached within 20 or 30 minutes, so almost pure juice is obtained for analysis, neutralization factors and gastric evacuation exert a minimum influence The test is usually performed the first thing in the morning with the subject fasting In a subject of average weight about 0.25 mg is injected

HYPOCHLORHYDRIA (SUBACIDITY) AND ACHLORHYDRIA (ANACIDITY)

When the stomach contents give values persistently below 20 "clinical units" (0.05 per cent) for free HCl after test meals, the condition is spoken of as *hypochlorhydria* or *subacidity* The complete absence of free HCl is referred to as *achlorhydria* or *anacidity* Bennett and Ryle in a study of 100 healthy male subjects (medical students) found achlorhydria present in 40 per cent. In the general run of hospital cases without gastric disease or pernicious anemia 14 to 20 per cent show anacidity It has already been pointed out that the absence of free hydrochloric acid from the stomach contents may be simply the result of excessive neutralization, and not of the suppression of acid secretion Again, there may be no secretion of acid after a test meal yet the glands respond to the more powerful stimulus of histamine, an acidity of this character is called *false* or *apparent anacidity* The failure of acid to appear after the injection of histamine is referred to as *true anacidity* The figures given above as reported by Bennett and

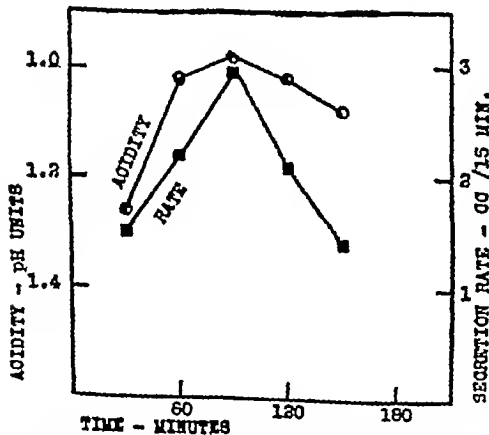


FIG 39 8 Curves of acidity and rate of secretion of the fundic pouch of a dog following a meal. (After Hollander and Cowgill)

Ryle for anacidity in healthy subjects as well as those given for hospital cases do not permit a distinction to be drawn between the two types of the condition, since test meals were used in the investigations. It is therefore impossible to say in what proportion of the cases true anacidity existed. The incidence of true anacidity appears, however, to be less than 1 per cent in young healthy male subjects and not more than 1 or 2 per cent in the case of young healthy females.

In cases of anacidity the peptic concentration of the gastric juice though usually low is in some cases not far below the normal value—further evidence that the secretion of acid and of pepsin are governed by independent mechanisms. Complete absence of both acid and pepsin from the gastric contents is called *achylia gastrica*.

The incidence of anacidity, both apparent and true, shows a definite increase with advancing years up to the sixth decade. Alvarez, Vanzant and associates, in an investigation of a large series of patients without gastric disease, found it present in from 25 to 35 per cent between the ages of 60 and 70 years. Females showed a higher incidence than males. After the sixth decade some reduction in the frequency of the condition was observed (see figs 39.6 and 39.9).

Anacidity in the majority of instances gives rise to no gastric symptoms and is compatible with perfect health. In some cases it is associated with flatulent dyspepsia and occasionally a persistent diarrhea (gastrogenous diarrhea) which is relieved by the administration of acid.

The following are the chief pathological conditions associated with anacidity,

(a) *Perniciosa anemia*. Achlorhydria is a constant phenomenon (p. 84). Pepsin is also usually absent. Histamine is ineffective.

(b) *Carcinoma of the stomach*. There is anacidity in over 60 per cent of cases. It is due to the chronic gastritis and atrophy of glandular tissue. This, it is believed, precedes by a variable period the development of the growth. The anacidity may be either false or true.

(c) *Chronic gastritis* causes a gradual depression of the secretory function and may lead finally to its complete suppression.

(d) Hypochlorhydria or achlorhydria may also occur during acute fevers, in malnutrition, gall-bladder disease, Addison's disease, sprue, acne rosacea and chronic arthritis.

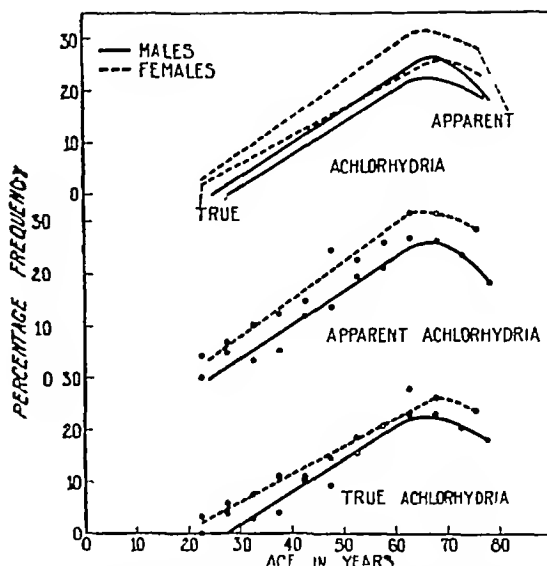


FIG. 39.9. Curves showing the relation between the incidence of achlorhydria and age (After Vanzant, Alvarez, and associates).

HYPERCHLORHYDRIA (OR HYPERACIDITY)

Hyperchlorhydria is also of common occurrence. About 5 per cent of healthy persons show the condition.

The free acid has a value of from 60 to 90 after a standard test meal, and may, instead of declining in the second hour, remain high or continue to rise. The high acidity of the gastric contents does not mean that the juice as secreted by the gastric glands is excessively acid. So far as is known the normal maximum of from 0.5 to 0.6 per cent of hydrochloric acid in the gastric secretion is never exceeded. The high acidity of the gastric contents is due either to the secretion of an abnormally large quantity of juice (hypersecretion) or to impairment of the factors regulating gastric acidity (p. 517), e.g., failure of the secretion rate to become reduced during the second hour, or delayed gastric evacuation.

Two pathological conditions are almost invariably accompanied by hyperacidity, namely *duodenal ulcer* and *pyloric obstruction* (non-malignant). In the last mentioned condition secretions and food materials accumulate in the stomach. The gastric contents are in consequence greatly increased, the stomach becomes dilated and large quantities of fluid are vomited from time to time. The loss of acid through vomiting may result in a condition of alkalosis accompanied by tetany (p. 849). The gastric distention occurring in this condition probably serves as the stimulus to secretion (p. 508).

The neutralization test This test is of value in investigating the acid regulating mechanism in duodenal ulcer especially in determining the success or otherwise of an operation performed for its cure. Instead of stimulating gastric secretion by a test meal or histamine injection, 300 cc of 0.5 per cent hydrochloric acid solution are introduced through a rubber tube, into the fasting stomach after its contents have been removed by aspiration. Samples of the acid solution are withdrawn from time to time as in the ordinary method of fractional gastric analysis, and the total acidities determined. The results are plotted and the curve compared with a normal standard curve. In the normal subject the total acidity falls from a value of 130 at the beginning of the test to around 40 within an hour or so. In cases of duodenal ulcer the total acidity falls more slowly and may be 70 or more after the lapse of 3 or 4 hours.

CHRONIC GASTRIC AND DUODENAL ULCER (PEPTIC ULCER)

Epigastric pain coming on usually in from a half to one and a half hours after a meal and vomiting are the chief clinical features of gastric ulcer. In a certain proportion (about 20 per cent) of cases blood appears in the vomitus (hematemesis).

In duodenal ulcer, pain occurs usually within from two to three hours after a meal, that is, when the stomach is nearly empty. The onset of the pain is therefore earlier after a light than after a heavy meal. The pain is relieved by taking food.

Pathogenesis

It is generally agreed that the dominant factor in the development of gastric and of duodenal ulcer is the action of the pepsin-hydrochloric acid of the

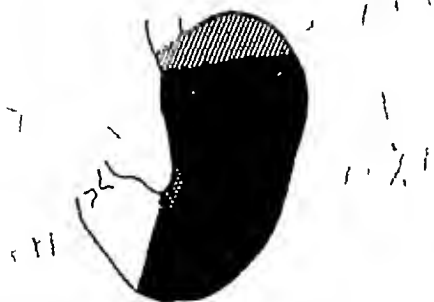


FIG 39 10 Diagram showing the distribution of the parietal (acid-secreting) cells in the human stomach. In the black area the proportion of parietal cells was maximal and was taken as 100 per cent, in the cross-hatched area on lesser curvature the percentage of parietal cells was 75 per cent, in the shaded area at the fundus, 50 per cent, and in the white area 0 to 1 per cent. (After Berger, Amer J Anat., 1934, 54, 87)

gastric juice. The term peptic ulcer is therefore well chosen. The importance of this factor is evidenced by the following facts.

(1) Apart from the ulcerations caused by some specific disease, e.g., tuberculosis, syphilis, carcinoma, etc., ulcer of the gastro-intestinal tract is confined to those regions which are exposed to the action of acid. (a) Gastric ulcers in the great majority of cases involve the pyloric part of the stomach, they are most frequently situated on the lesser curvature near the incisura angularis or on the anterior or posterior wall in close proximity to this limited area. They are never seen in the dome of the fundus and rarely on the upper part of the greater curvature, these regions, it will be noted, are not in contact with acid for any length of time. (b) The lower part of the esophagus into which highly acid juice frequently regurgitates, and the adjacent part of the stomach wall, i.e., the cardia, are sometimes the site of ulceration. (c) Duodenal ulcer occurs practically exclusively within the first inch or less of the duodenal cap (p 570), and nearly always upon its anterior or posterior wall, that is, where the chyme before it has been neutralized by the alkaline juices of the duodenum comes into contact with the mucosa. (d) After gastrojejunostomy, the so-called stomal ulcer may occur in the jejunal mucosa in the region of the anastomosis, i.e., where the gastric juice first impinges. (e) In a Meckel's diverticulum which contains ectopic gastric glands an ulcer occasionally forms. The ulcer's site is either in that part of the mucosa of the diverticulum, which does not itself contain acid-secreting glands, or in the ileum at the point where the diverticulum opens into it. Matthews and Dragstedt, experimenting with dogs, produced an "artificial Meckel's diverticulum" by transplanting a pouch of the gastric wall into the ileum, an ulcer developed in the ileum just beyond the transplant in every experiment. These observations emphasize a curious fact that the commonest situations of ulcer are not in the mucosa which itself secretes the acid, but in neighboring parts which normally secrete a neutral or alkaline fluid—the pyloric region, duodenal cap, cardiac region, esophagus, jejunum or ileum. The pyloric type of gland extends farther up the lesser than up the greater curvature (fig 39 10). It has even been suggested that the occurrence of ulcers in the body of the stomach, i.e., in the acid-secreting part of the mucosa, is actually dependent upon the presence of patches of aberrant pyloric glands. (f) Mann and Williamson, employing dogs, excised

the duodenum and transplanted it into the ileum, thus diverting its alkaline juices away from the region of the pylorus. The cut end of the jejunum was anastomosed to the pylorus. Fourteen out of sixteen animals upon which this operation was performed developed chronic ulcers in the jejunum just beyond the pylorus. However, Fawley and Ivy found that this operation, if combined with excision of the fundus, or if the alkaline secretions are drained into the stomach, does not cause a jejunal ulcer.

(2) In a very large proportion of subjects of duodenal ulcer the concentration of free hydrochloric acid in the gastric contents after a test meal is abnormally high. The interdigestive or basal secretion especially during sleep is increased much above the normal, the pepsin concentration of the secretion is also frequently increased. According to Hurst, hyperchlorhydria (p. 519) is present in 61 per cent of cases. Of the remainder the majority show an acidity which is near the higher limit of the normal, a few only have subacidity. Though a small acute or subacute ulcer may sometimes develop in the absence of acid, a large chronic ulcer of the duodenum is almost never seen with anacidity, as in pernicious anemia. In duodenal ulcer the typical findings upon gastric analysis are:—a fasting juice of greater volume than normal and of high acidity, and a curve of gastric acidity after a test meal which rises well above (20 units or more) the normal maximum. In some cases the stomach empties more rapidly than usual as a result of exaggerated gastric motility and the curve of acidity

after reaching its maximum value falls steeply again. In other instances the stomach empties in the usual time but the curve of acidity is maintained at its maximum as a result of continued gastric secretion—the *plateau type of curve*. In still other instances the emptying of the stomach is delayed as a result of pylorospasm or achalasia (see below), and the curve of gastric acidity instead of falling at the usual time continues to rise—the so-called *climbing curve* (fig. 39.11). High gastric acidity is less commonly associated with gastric ulcer, according to Vanzant the acidity is actually a little below the normal average. True anacidity, however, is rarely, if ever, found. The relatively low gastric acidities found in gastric ulcer are probably the result of an associated gastritis and do not necessarily indicate that hyperchlorhydria did not precede the development of the ulcer.

(3) Stimulation of gastric secretion, as by the continued administration of histamine or of caffeine (p. 518), is one of the most effective experimental means of producing gastric ulcer. Although gastric ulcers are produced experimentally by means of a constant drip over the gastric mucosa of a solution of hydrochloric acid, some pepsin must be furnished by the stomach itself. Irrigation of a loop of jejunum with acid alone will not cause ulceration but this is readily produced by the addition of pepsin to the perfusion fluid.

(4) Measures directed toward the prevention of excessive gastric secretion and toward the neutralization of the acidity of the gastric contents are

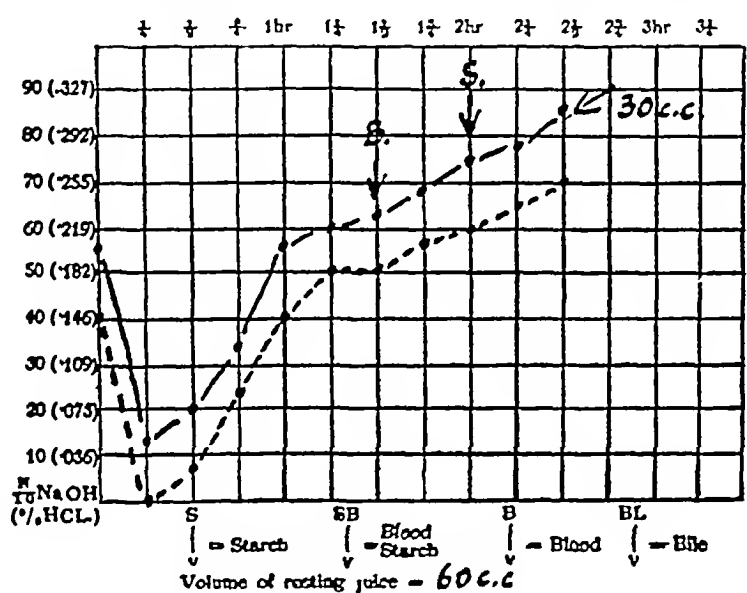


FIG. 39.11 Chart showing "climbing" type of curve of gastric acidity (After Ryle)

of outstanding value in encouraging the healing of the ulcer (p 525)

Though the importance of acid in the production of ulcer cannot be denied this factor cannot be solely concerned For one thing, many persons who show hyperchlorhydria do not develop ulcer "Why in these instances is the gastric mucosa immune to the action of the pepsin-hydrochloric acid?" Indeed, the question has often been asked, "Why does not the pepsin-hydrochloric acid of even normal gastric juice digest the gastric or duodenal mucosa?" It is also an extraordinary fact, firmly established by several workers, that the tissue of other parts of the intestinal tract or of other organs, though susceptible to the action of acid in their normal situations, are not digested when transplanted into the wall of the normal stomach Dragstedt and Vaughn, for example, removed areas from the gastric wall of dogs and then sutured portions of the duodenum, ileum, jejunum, colon, spleen or kidney into the gaps In no case was the transplanted tissue digested In the case of the kidney and spleen their gastric surfaces became covered with a layer of gastric epithelium Sections of transplanted intestinal mucosa were found to be perfectly normal after a period of nine months

On the other hand, as shown years ago by Claude Bernard, the intact leg of a living frog is digested when introduced through a fistula into the stomach of a dog Pavy showed the same thing for the rabbit's ear and Dragstedt and Vaughn have demonstrated that the intact limb of a live frog placed in an extract of frog's gastric mucosa is digested

Until comparatively recently there has been no convincing or even very plausible answer to the question of why the stomach does not digest its own wall One explanation offered was that the greater alkalinity of the blood coursing through the vessels of the gastric tubules as a result of the loss of H ions served to neutralize the effect of the acid juice Others suggested that the immunity of the gastric mucosa to autodigestion is due to its containing an antipepsin Another suggestion is that the mucus which coats the mucosa exerts a protective action The probable explanation is to be found in the production of ammonia in the gastric wall by the action of urease on urea

UREA-UREASE AMMONIA MECHANISM. The presence of ammonia in the gastric contents was observed many years ago (1852) by Bidder and

Schmidt, but little significance was attached to the finding until Luck and Seth demonstrated its production in gastric tissue by the action of urease upon urea ($\text{urea} + \text{urease} \rightarrow \text{NH}_3 + \text{CO}_2$), and pointed out the possible role of this mechanism in neutralizing gastric acid A rise in blood urea as in renal insufficiency, or after the oral administration, of urea is followed by a corresponding increase in the urea + ammonia in the gastric wall, and a rise in the ammonia concentration of the gastric juice In the gastric mucosa the concentration of urease is higher than in any other tissue, it is found in greatest amounts in the cells of the surface epithelium Fitzgerald and his associates have made an extended study of this mechanism Hastings and his colleagues found that the total urea content of the body (as determined by urea tagged with C^{14} and an analysis of the expired air for C^{14}O_2) was 10 per cent per hour, this was reduced after gastrectomy

A CONSIDERATION OF OTHER FACTORS CONCERNED IN THE PRODUCTION OF ULCER *Bacterial infection*, and interference with the *blood supply* to the mucosa, either as a result of emboli or thrombosis have been thought by some to be responsible for ulcer production Except perhaps in rare instances, these are no longer believed to play a rôle *Tobacco smoking* has been thought in some instances at any rate, to encourage ulcer formation or to interfere with the healing of an ulcer already formed But Schnedorff and Ivy found that smoking had a depressing effect, if any, upon gastric acidity, and evacuation of the stomach was delayed Both of these effects should be beneficial rather than otherwise and if smoking is conducive to gastric or duodenal ulceration the manner in which it acts is not clear *Trauma*, although not essential to the production of ulcer is probably often a contributory factor It is not difficult to believe that in the presence of other causative factors, rubbing of food against the gastric mucosa or the passage of coarse indigestible material into the duodenal cap, will encourage the production of ulcer, or retard the healing of an existing one It is to be remembered that food after entering the stomach passes along the lesser curvature, i.e., through the *magenstrasse* (p 564), and, as already mentioned, this region is one of the commonest sites of gastric ulcer Experimental erosions in this situation heal much more slowly than do those produced along the greater curvature. Mann and Bollman also point out that the site of duodenal

ulcer corresponds to the area of mucosa upon which the gastric contents impinge, and, when the gastric movements are energetic, this may occur with considerable force. They found that the experimental production of ulcer was considerably delayed if the propulsive force of the stomach was reduced by making an hour-glass constriction in the prepyloric region. Ivy and his associates have also shown that in the rabbit coarse food retards the healing of an area of the stomach wall from which the mucosa has been denuded.

The possible effect of trauma upon the blood supply of the gastric mucosa is referred to on page 522.

The neurogenic element in the production of ulcer

Cushing was the first within recent years to stress the importance of nervous influences in the pathogenesis of ulcer. He drew attention to the relatively high incidence of acute gastric ulcers after intracranial operations. Others previously had remarked upon the association of tumors of the mid-brain and diencephalon with gastric or duodenal ulcers, and the older pathologists had mentioned the frequency of softening of the gastric wall (gastromalacia) and other gastric lesions in subjects dying of cerebral conditions. Cushing suggests that the influences arising in the parasympathetic center in the hypothalamus (p. 1024) and conveyed along the vagus nerves, are responsible for changes in the gastric mucosa which lead to the development of ulcer.

The experimental investigations of a number of workers support this conception. Beattie, for example, produced areas of hyperemia and small erosions in the gastric mucosa (lesser curvature) by stimulation of the hypothalamus in the region of the tuber cinereum. Keller and his associates have observed ulcerations of the stomach and proximal duodenum with small hemorrhages, following lesions of the hypothalamus. Hyperemia and erosions of the gastric mucosa have also been produced by the injection of pilocarpine (a parasympathetic stimulant) into the third ventricle or by continued stimulation of the vagus nerve. It may also be recalled that stimulation of the hypothalamus in the region of the tuber cinereum is followed by movements of the stomach (p. 570), and that hypersecretion and gastric hypertonus (steer-horn type of stomach) are prominent features of duodenal ulcer. Moreover, it is recognized clinically that the nervous "highly strung" type of

person is more likely to be the subject of ulcer than is the phlegmatic type, and that an ulcer in the process of healing not uncommonly relapses as a result of some nervous influence, worry, emotional shock, concentrated mental work, etc. The profound effect of psychic states upon the vascularity of the gastric mucosa has been described. Babkin suggests the possibility that vagal impulses cause the liberation of histamine in the gastric mucosa. Through its vasodilator action and stimulating effect upon the parietal cells, conditions favorable to ulcer production are provided, namely, high gastric acidity and, through capillary stasis, defective blood supply to the mucous membrane.

It may well be as Gray and his associates suggest, that the effects of stress upon the development of ulcer are due to hormonal rather than to direct nervous influences, namely through ACTH and the discharge of cortisone from the adrenal cortex. In normal subjects either of these principles increases the concentration of acid and pepsin in the gastric juice, and in patients with duodenal ulcer hemorrhage or impending perforation was induced by their administration.

A possible mode of production of gastric ulcer through nervous impulses is suggested by the discovery by Barclay and Bentley of arteriovenous anastomoses in the gastric submucosa. In stomachs excised for duodenal ulcer the finer mucosal vessels did not admit injection material, due apparently to the opening up of larger channels through which the blood had been shunted. Ischemia of the mucosa induced by shunting of blood through deeper vessels could quite obviously lead to nutritional changes, lowered vitality of the tissue and its digestion by the gastric juice. Operative trauma appeared to be the cause, through a spinal reflex, of the redistribution of blood through these shunts, for it was prevented by spinal anesthesia. But it is easy to conceive of a similar effect being induced by impulses from higher nervous centers.

The experimental production of gastric ulcer. Besides the methods already mentioned, namely, by diverting the alkaline secretion into the ileum and stimulation of the hypothalamic region, gastric ulceration can be produced in animals by the daily oral administration of *cinchophen* or by the injection of *pitressin*. Gastric ulcers can also be produced by the continued action of *histamine* or *caffeine*. They are administered intramuscularly in beeswax in order to prolong their stimulating effect upon gastric secretion. The mode of action of *cinchophen*

(which also in large doses causes liver damage) is unknown. Pitressin (or pituitrin) apparently induces superficial necrosis of the mucosa as a result of vascular spasm (see Dodds and associates). Byrom has shown that following pituitrin administration necrotic areas accompanied by small hemorrhages occur in other organs as well.

The cause of the pain in duodenal ulcer

As already mentioned the pain of duodenal ulcer commences when the stomach is nearly empty, that is, usually about two hours after a meal. There is not general agreement as to the mechanism whereby the pain is produced. It is generally conceded, however, that it is not due to the direct stimulation of pain endings in the ulcer by acid or by food, pain fibers of the gastro-intestinal tract are not sensitive to chemical or the ordinary types of mechanical stimuli, tension is their adequate stimulus (p. 599). Nevertheless, that the acidity of the chyme is an important factor in the production of pain is evidenced by the experiments of Palmer and others. Palmer found that the introduction of 300 cc. or so of a 0.5 per cent solution of hydrochloric acid into the empty stomach of a large proportion of ulcer patients caused pain. This procedure aroused no sensation in normal persons. The pain is therefore considered by most observers to be a consequence, though an *indirect* one, of the stimulation of afferent nerves in the base of the ulcer by the acid chyme. Disturbances of the motor mechanism of the stomach or duodenal cap (p. 570) dependent upon reflexes initiated from the irritable focus are believed to be the cause of the pain.

An observation of Dragstedt and Palmer, however, suggests that in some instances the nerves of the ulcer, as a result of inflammation, may have become so hypersensitive that they give rise to pain in response to stimuli of a type to which healthy visceral nerves are insensitive. It was found that gentle stroking of the surface of an ulcer, in a conscious patient during an operation, caused pain. Although spasm of the intestinal muscle occurred in the region of the ulcer, and this was painful, the ulcer surface itself appeared to be directly sensitive to mechanical and chemical stimulation. Irrigating it with acid was followed by pain which was relieved by the application of sodium bicarbonate. Evidence has come from other quarters that the gastric nerves subserving pain, which normally are insensitive to the ordinary types of stimulus, may be so altered by inflam-

mation as to give rise to pain when stimulated by acid. Jones found that whereas HCl, alcohol, sodium hydroxide, and other chemicals when applied to the interior of the healthy human stomach aroused no sensation, severe pain was caused when an inflamed mucosa was treated similarly. Bonney and Pickering from experiments upon patients with duodenal ulcer showed that emptying the stomach brought relief, but the return of the acid fluid to the stomach caused the pain to return. They believe that the pain was due to the direct stimulation of the abnormally sensitive nerves by acid.

Although motor disturbances of the pyloric and duodenal region are probably, in most instances, responsible for the ulcer pain, there is not general agreement as to their precise nature. According to some, spasm of the pylorus (pylorospasm), is responsible. Hurst considers that *achalasia* of the sphincter, i.e., a failure to relax upon the approach of a peristaltic wave (p. 562) rather than actual spasm exists. The immediate cause of the pain he believes to be the distension of the wall of the pyloric antrum as a powerful peristaltic wave drives fluid downwards against a pylorus which fails to open. The observations that pain may occur in the absence of peristaltic activity in the stomach, and be absent during gastric hypermotility, and that the pain is often continuous rather than intermittent, are opposed to this conception. Ryle attributes the pain to a general increase of gastric tone. The researches of Wilson indicate that the pain of duodenal ulcer is due to contraction of the duodenal cap. He examined the stomach of patients with duodenal ulcer under the fluoroscope after the ingestion of a barium meal, and during the pain. By moderate pressure upon the stomach through the abdominal wall he was able to drive fluid through the pylorus. In 14 out of 16 subjects the entrance of fluid into the duodenal cap was followed by relaxation (receptive relaxation) of the latter and the relief of pain. The ease with which material was pressed into the cap argues against the existence of a degree of pylorospasm, or *achalasia* which could result in painful distention of the antrum. Furthermore, in many instances pain was experienced though the stomach was free from peristalsis. In view of these observations associating contractions of the wall of the duodenal cap with ulcer pain, it is quite conceivable that such contractions by causing tension to be exerted upon inflamed nerve endings in the base of the ulcer are responsible for the sensation.

On the other hand, if the pain originates in this way how are we to account for the immediate relief which frequently follows the ingestion of food or of sodium bicarbonate? The relief, in some instances, occurs too rapidly to be accounted for simply by a reduction in acidity of the gastric contents. Are we to assume that in the case of bicarbonate ingestion the gas evolved passes at once through the pylorus and causes relaxation of the walls of the duodenal cap in the same manner as does fluid material forced through the pylorus by manual pressure? On the other hand, does the ingested food, or the evolved gas, by increasing the volume of the gastric contents abolish the pain by causing an adaptive relaxation of the gastric walls (pp 557 and 564). This is Ryle's suggestion. Definite answers to these questions are not forthcoming and the immediate cause of the pain of duodenal ulcer is still under debate.

Pain, in some instances, is due, apparently, to the spread of the inflammatory process to the serous coat and the traction of peritoneal adhesions. Pain of this nature is influenced by posture, disappearing usually when the subject reclines.

A new explanation of ulcer pain is offered by Kinsella, namely, engorgement of the vessels in the indurated base of the ulcer and the resulting increase in tissue tension. At the time that the pain comes on the stomach is emptying, then, according to this vascular conception, the capillary bed of the gastric wall is compressed and blood forced into the less resilient ulcer area. The relief from pain after food is attributed to the withdrawal of blood from the ulcer site to other parts of the gastric mucosa. Kinsella believes that acid and the motility of the stomach play secondary rôles in the production of ulcer pain. Thus, the pain of ulcer is placed in the same category with pain caused by a chronic inflammatory lesion in other situations, e g, in a boil or an ulcer of the leg. It is well known that the pain associated with inflammation is aggravated by hyperemia.

The application of physiological principles to treatment

The modern medical treatment of duodenal ulcer comprises the following measures: (a) Meals at frequent intervals, this practice has the two-fold effect of reducing gastric acidity and furnishing digestive products (probably peptones) which, as shown by the experiments of Mann and Bollman, exert a protective action against the action of acid. (b) The inclusion in the diet of such materials as

cream and milk which, by inhibiting gastric secretion and motility, provoking the expulsion of bile from the gallbladder and stimulating the secretion of pancreatic juice, tend to reduce the acidity of the duodenal contents. (c) The limitation of highly seasoned foods and materials rich in extractives, especially meats and alcohol, which stimulate the gastric glands. (d) The omission from the diet of coarse indigestible articles, e g, raw fruits and vegetables, which are likely to cause mechanical irritation of the ulcer. (e) The administration of alkalis, such as sodium bicarbonate to neutralize the acid, or of mucin which has a high acid-combining power (p 502), and serves to form a protective coating for the ulcer. Enterogastrone (p 512) has been advocated as a means of inhibiting gastric secretion and motility. There is real danger of inducing alkalosis by the prolonged administration of readily absorbable alkaline salts such as sodium bicarbonate. Tri-calcium phosphate, calcium carbonate and magnesium oxide, and colloidal aluminum hydroxide although less efficient antacids are much less likely to induce alkalosis owing to their relatively slight absorption. The last mentioned chemical is said to inhibit secretion and to inactivate pepsin, but these substances, in general, are merely neutralizing agents and do not affect the secretion of acid. Some, such as sodium bicarbonate, may actually increase secretion. Synthetic anion exchange resins (e g, Resinat) used for the same purpose, have a high neutralizing power, inhibit peptic activity and are non-absorbable. (f) Atropine is sometimes employed to reduce secretion. (g) On the basis of the urea-urease-ammonia mechanism urea administration has been employed with success in the treatment of ulcer. (h) The avoidance of overwork, mental or physical fatigue, worry or any psychic state likely to bring the neurogenic factor into play (p 523).

In the treatment of duodenal ulcer, surgery is resorted to much less frequently today than in the past. When some complicating feature (e g, hemorrhage, perforation or pyloric obstruction) calls for operative treatment, one or other of the following procedures is usually undertaken with the object of effecting a radical cure: (a) Excision of the ulcer and reconstruction of the duodeno-pyloric region (pyloroplasty). (b) Anastomosis of the jejunum to the posterior aspect of the stomach at the lower part of the greater curvature (posterior gastroenterostomy). (c) A combination of (a) and

(b) (d) Resection of the pyloric region of the stomach and of the ulcer-bearing area, and then either (i) anastomosing the cut end of the duodenum directly to the gastric stump (Billroth I type of operation), (u) closing the duodenum and the gastric stump and performing a posterior gastro-jejunostomy (Billroth II operation) or (m) closing the duodenum and joining the jejunum to the gastric stump (Polya-Balfour operation)

The operation of supradiaphragmatic section of the vagus nerves (vagotomy) introduced by Dragstedt and associates in 1944 has been practiced extensively for gastric and duodenal ulcer. Its aim is the reduction of the continuous (interdigestive) secretion of gastric acid (p. 505), especially during the night. The gastric and intestinal phases of secretion are little affected by this operation.

CHAPTER 40

DIGESTION IN THE INTESTINE

THE PANCREAS

STRUCTURE The pancreas is a racemose gland, its alveoli resembling those of the salivary gland in their general arrangement and design. Lying between the alveoli are groups of cells constituting the Islands of Langerhans which furnish the internal secretion of the pancreas (insulin, ch 49). The cells which line the alveoli and furnish the ferments of the pancreatic juice, contain, like the serous cells of the salivary glands and the chief cells of gastric tubules, zymogen granules which become reduced in number during secretory activity, and increase again after a period of rest. Two distinct zones may be seen in these cells, an inner lying next the alveolar lumen and containing the zymogen granules, and an outer which is clear and homogeneous. During secretion the outer clear zone increases in depth at the expense of the granular zone.

The secretion is collected by a branching system of ducts which eventually lead into the main pancreatic duct—*duct of Wirsung*. This discharges in company with the common bile duct into the ampulla of Vater, which opens through a small orifice, situated upon the summit of a small papilla about $3\frac{1}{2}$ inches below the pylorus. In a certain proportion of cases the pancreatic and common bile ducts open into the intestine by separate orifices. An accessory pancreatic duct—*duct of Santorini*—is usually present. It opens into the duodenum about $\frac{1}{4}$ inch above the duct of Wirsung, it also sends a branch to the latter through which it delivers the greater part of its contents.

The acinar and islet tissues are supplied by separate arterioles. The gland receives a rich innervation from both sympathetic and parasympathetic sources.

THE COMPOSITION OF PANCREATIC JUICE

The composition of pancreatic juice is given in table 37 modified from Starling. It is alkaline in reaction, due mainly to a high content of sodium bicarbonate, whose concentration is from 100–150 milliequivalents per liter. Its pH is between 7.10 and 8.20 (in the dog).

The ferments of pancreatic juice are *trypsin*, *chymotrypsin*, *amylase*, *lipase*, *rennin* and *maltase*. Unlike the gastric enzymes, apparently those of pancreatic juice are produced by one type of cell

Pancreatic proteinases

There are at least two proteinases in pancreatic juice—*trypsin* and *chymotrypsin*. Trypsin is not

secreted as such but in an inactive form called *trypsinogen*. The activation was first shown by Schepowalnikow in Pavlov's laboratory to be brought about by the addition to pancreatic juice of a small quantity of intestinal juice or an extract of the intestinal mucosa. The activation is due to an enzyme called *enterokinase*. It has been denied by some (e.g. Waldschmidt-Leitz) that enterokinase is a true enzyme, but, as a result of the work of Kunitz, its enzymic nature is now accepted. Trypsinogen undergoes spontaneous transformation to trypsin when the pancreas is allowed to stand in a slightly acid solution. Pure trypsinogen is also activated by the addition of trypsin, or of magnesium or ammonium sulphate. The spontaneous (autocatalytic) activation of trypsin is greatly accelerated by the presence of a calcium salt.

Chymotrypsin, like trypsin, is secreted in an inactive form, *chymotrypsinogen*, and is activated by trypsin but not by enterokinase. These two enzymes, trypsin and chymotrypsin, are mainly responsible for the proteolytic activity of pancreatic juice. Both enzymes are proteins and they, as well as their precursors, have been obtained in pure crystalline form. Trypsin has no milk-curdling power, but is capable of clotting blood, whereas chymotrypsin, although it will not cause coagulation of blood has a powerful milk-curdling action, and is probably identical with pancreatic rennin. The optimum pH for the action of trypsin and of chymotrypsin is around 8, that of enterokinase is within the range between pH 5.2 to pH 6.0.

The *action* of trypsin carries the digestion of protein beyond the peptone stage. It also differs from peptic digestion in being carried on in an alkaline instead of in an acid medium. It is possible, moreover, that different linkages in the protein molecule are attacked by the two enzymes. Though under ordinary circumstances trypsin is called upon to commence its action after the gastric juice has converted a large part of the protein into proteoses and peptones, it can also attack native protein, flesh introduced directly into the duodenum is readily digested by the pancreatic juice. Nevertheless, the preparatory digestion of protein by gastric juice is favorable to tryptic action. The specific action of trypsin is to break the proteose

Disorders of the digestive function of the pancreas

In chronic pancreatic disease (chronic pancreatitis) the digestive power of the juice secreted by the gland becomes markedly reduced with the result that the feces contain quantities of undigested fat, protein and starch. The stools are increased in bulk and pale, and there may be diarrhea (pancreatogenous diarrhea). Several clinical tests have been devised in order to determine the extent of the impairment of the digestive functions of the gland. The microscopical examination of the feces for intact muscle fibers after a meal of meat, or the estimation of the fecal fat often gives valuable information. Others estimate the enzyme activity of the duodenal contents withdrawn through a duodenal tube. Acute pancreatitis is associated with severe abdominal pain, vomiting and symptoms of shock. In the more severe types of the acute disease hemorrhages into the glandular substances with necrosis occur. Fat necrosis in the mesentery, omentum and peritoneum, as well as in the pancreas itself is commonly observed. This is caused by the escape of lipase from the disorganized gland.

Hemorrhagic pancreatitis has been thought to be the result of the obstruction of the outlet of the ampulla of Vater (gallstone, spasm of sphincter of Oddi or swelling of duodenal mucosa) and the passage, in consequence, of bile into the pancreatic duct system (Archibald). However, according to Dragstedt and his associates only about 60 per cent of cases of hemorrhagic pancreatitis arise in subjects of chronic biliary tract disease and in only about 10 per cent of these is the ampulla obstructed by a gallstone. Rich and Duff find that the reflux of bile into the pancreatic ducts is a rather infrequent accompaniment of hemorrhagic pancreatitis. When reflux of bile is observed obstruction of the pancreatic duct system with subsequent rupture of the duct wall and the escape of pancreatic juice into the interstitial tissue of the pancreas is believed by these observers to be the essential factor in the production of the hemorrhagic lesion, rather than the action of the bile itself. Obstruction to the flow of pancreatic juice, due to metaplasia of the epithelial lining of a branch of the pancreatic duct within the gland, is considered by Rich and Duff to be the commonest cause of acute pancreatitis. The ductules and acini behind the obstruction become dilated, rupture of the distended walls and the escape of digestive enzymes into the glandular substance is looked upon as the immediate cause of the disease. The interstitial tissue of the pancreas, in common with other tissues of the body, possesses the power to activate the trypsinogen of the pancreatic juice, digestion and necrosis of the blood vessels result. These observers found metaplasia of the duct epithelium and acinar dilatation in 13 of 24 cases of hemorrhagic pancreatitis. In most cases of the disease a gallstone was not found in the ampulla of Vater and the main pancreatic duct was unobstructed. It is pointed

out that rupture of a dilated ductule or acinus is most likely to occur when the secretion pressure of pancreatic juice is high, i.e., after a large meal or the ingestion of alcohol.

A determination of the concentration of the amylase or lipase of the serum or of the output of amylase in the urine is valuable as a means of recognizing acute pancreatic disease. In pancreatic damage the urinary amylase and the concentration of amylase and of lipase in the serum are greatly increased. The *secretin test* is of great value in investigation of pancreatic insufficiency, especially in chronic disease, e.g., cystic fibrosis of the pancreas. In performing the test a purified preparation of secretin is injected intravenously and the duodenal secretions withdrawn by suction through a duodenal tube. The secretions are collected in 20 minute periods for an hour following the injection and the enzyme activity (trypsin, lipase, amylase, and phosphatase) determined.

THE EFFECTS OF THE TOTAL DRAINAGE OF PANCREATIC JUICE TO THE EXTERIOR. Elman and McCaughan have demonstrated the fatal effect of the loss from the body of all the pancreatic juice. In dogs the survival time following the establishment of a pancreatic fistula through which the juice was drained to the exterior was from seven to eight days. Loss of electrolytes with consequent dehydration (p. 24) is undoubtedly a factor leading to this result, for the life of the animals can be prolonged by the intravenous administration of Ringer's solution. Death cannot, however, be prevented by this treatment, the animals succumbing after some weeks in spite of it. Nor do the animals die simply as a result of inanition, for it has been shown that in the absence from the intestine of the pancreatic juice either as a result of a pancreatic fistula or of pancreatectomy, sufficient quantities of food are digested and absorbed to maintain life for indefinite periods. It has been suggested, therefore, that the loss of some material essential to life other than salts and water is the cause of death.

Ligation of the pancreatic ducts in dogs is followed by defective digestion and absorption of all three types of food stuff. There is pronounced polyphagia, loss in weight and impairment of liver function (Ivy). The general nutrition of animals is improved by feeding raw pancreas. Similar defects of digestion occur in man when the flow of pancreatic juice is obstructed. But the depression of fat digestion and absorption varies considerably in animals and man, the range being from 0 to 92 per cent of fat absorbed. There is no satisfactory explanation for this variability.

THE SECRETION OF PANCREATIC JUICE

Small quantities of pancreatic juice are secreted continuously. Secretion is under both hormonal and nervous control.

The hormonal control

In 1902 Bayliss and Starling showed that an acid (HCl) extract of the duodenal mucosa when injected into the blood stream of an animal caused a copious flow of pancreatic juice. Intravenous injection of the acid itself was ineffective. A secretagogue effect also followed the introduction of acid into a loop of bowel whose nervous connections had been completely severed, the only communication between the bowel and the pancreas being then through the blood stream. The secretory effect was shown to be specific and not simply due to vasodilatation of the pancreatic vessels. The excitatory substance was called *secretin*.

The obvious physiological implication of these results was that the acid chyme upon coming into contact with the duodenal mucosa caused the production, or liberation, of a substance which was then conveyed by the blood to the pancreatic cells. Ivy, Farrell and Lueth transplanted a loop of intestine and the tail of the pancreas to subcutaneous sites, thus isolating the two structures indubitably from extrinsic nervous control. They observed secretion from the pancreatic transplant when acid was placed in the isolated loop. Since the intravenous injection of acid alone is ineffective, this experiment affords conclusive evidence for the liberation of a hormone from the intestinal mucosa.² Mellanby and Huggett in 1926 demonstrated that secretin existed preformed in the mucosa from which it could be extracted by water, alcohol and other solvents as well as by acid. Bayliss and Starling had claimed that the hormone existed in an inactive or precursory form which they called *prosecretin*. Mellanby also found that bile introduced into the duodenum caused secretin to be absorbed into the blood stream and that the active agent in the bile was the cholic acid of the bile salts. He suggested that the bile salts in their passage through the intestinal mucosa absorbed the secretin and carried it into the blood. Bile salts, however, though aiding in the absorption of the hormone are not absolutely essential. Ivy and his colleagues have shown that food (meat and fat) entering the intestine stimulates pancreatic secretion in the usual manner after ligation of the common bile duct.

Secretin exerts a hydrelatic action,³ that is, it

² The word, hormone, was used for the first time as a result of these experiments. They also furnished the first conclusive evidence of a hormonal mechanism.

³ *Hydrelatic* and *ecbolic* are terms introduced by Babin to indicate, respectively, the secretion of water or of enzymes.

causes the secretion mainly of water and the inorganic constituents of the juice (see below). Its action is little affected by atropine in moderate dosage, or by ergotamine which paralyzes sympathetic secretory fibers, so there is no evidence that its production or absorption is influenced by nervous mechanisms, or that it stimulates parasympathetic or sympathetic terminals in the pancreas. It appears to be a direct excitant of the glandular cells. Secretin also stimulates the secretion of bile and probably of the succus entericus.

An influence of the hypophysis on the secretin content of the small intestine has been demonstrated by Dorchester and Haist. In rats, the extractable hormone is significantly reduced after hypophysectomy, but this result could be prevented by daily injections of an anterior lobe extract, by ACTH, or by the growth hormone. A slight increase in the secretin content was induced by injections of an anterior lobe extract into normal rats, but not by ACTH or the growth hormone.

Agren and associates have prepared secretin in crystalline form. It is a polypeptid (molecular wt. 5000) giving a positive biuret but a negative ninhydrin reaction. It is basic in character, containing histidine and arginine but is free from cystine, tyrosine and tryptophane. This preparation and a cruder but non-toxic amorphous material has been used by these workers to test biliary and pancreatic functions in man. Ivy and Greengard have obtained secretin in the form of a crystalline picrolonate. The highest yield of secretin is given by the upper two-thirds of the small intestine. Minimal amounts are obtained from the lower third of the small intestine and from the ascending colon. It is absent from the gastric mucosa.

Pancreozymin This is the name which Harper and Raper have given to a second pancreatic excitant obtained, like secretin, from extracts of duodenal mucosa. Its existence has been confirmed by Greengard and his associates. Unlike secretin it stimulates the secretion of trypsin, amylase and lipase (*ecbolic effect*) but exerts little or no effect upon the volume of secretion. It causes the discharge of the zymogen granules, its action is thus similar to that of the vagus, though its effect is apparently a direct one upon the gland cells and not exerted through the nerve terminals, for it is not altered by atropine.

It cannot be definitely stated that pancreozymin is a hormone for it has not been shown to play a physiological role, i.e., discharged into the blood stream during digestion, as has been proved for secretin.

The control of pancreatic secretion by nerves

It has been known since the important work of Pavlov on pancreatic secretion that the gland cells are excited by vagal impulses. Control of pancreatic secretion by the vagus was shown in the following way. The nerve was divided in the neck four days previously. This preliminary procedure has the effect of abolishing the irritability of the cardiac fibers since they degenerate before the secretory fibers. When the nerve prepared in this way is stimulated a secretion of juice occurs, but it is much less obvious than that caused by secretin or may not be evident at all. This led to the belief that nervous control of pancreatic secretion was of little importance, or that the vagus contained inhibitory as well as the excitatory fibers, and that the effect of the former overshadowed the effect of the latter. But there are two reasons for the inconspicuous effect of vagal stimulation. First, vagal secretion as compared with the secretion evoked by secretin is viscous and of small volume, and, secondly, a coincident contraction of the larger ducts results, which blocks the flow. The vagal secretion is particularly rich in ferments (ecbohic effect) and is accompanied by a reduction in the zymogen granules of the cells. The chief effect of secretin, on the other hand, is to cause the secretion of the water and inorganic constituents, e.g., the bicarbonate of the pancreatic juice. Only after repeated injections of secretin does a reduction in the zymogen granules occur, and this effect is always less pronounced than that following vagal stimulation. Apparently then, the extrusion of the colloidal zymogenous particles from the cells into the alveoli is largely controlled by nervous impulses. Secretin, on the other hand, causes a flow of alkaline fluid which serves to flush the alveoli, to thin the juice rich in organic materials and sweep it along the ducts. Pilocarpine acts similarly to vagal stimulation. Atropine annuls the nervous secretion but, as already mentioned, has little effect or none at all on the action of secretin or of pancreozymin. Stimulation of the splanchnic nerves was found by Harper and Vass to reduce the rate of secretion of pancreatic juice and the output of enzymes, but this may be secondary to a constrictor effect upon the vessels of the gland. They conclude that the sympathetic is either without effect upon pancreatic secretion or is inhibitory.

Pavlov showed that there is a psychic element in the control of pancreatic secretion. Though the secretion in this cephalic phase is much less than in the case of gastric secretion, sham feeding was

found very definitely to increase the pancreatic flow—a result which Ivy has confirmed.

THE ACTION OF THE CHYME UPON PANCREATIC SECRETION

Chyme escaping through the pylorus causes an approximately equivalent amount of alkaline fluid (bile and pancreatic juice) to enter the intestine. Among the substances in the chyme which exert the greatest secretagogue effect upon the pancreas are acids, meat extracts, protein derivatives, fats, fatty acids, soaps and water.

There is evidence that peptones, acids and soaps exert their action through local reflex mechanisms and that these in turn are influenced by vagal impulses. Crider and Thomas found that the effect of peptones and of acid (but not of soap) upon secretion was abolished immediately after section of the vagus nerves. Recovery occurred a few days later but the juice was reduced in volume and in its concentration of total nitrogen. They suggest that the vagus exerts its control upon the pancreatic secretion not directly but through augmentation or inhibition of the activity of the local reflexes.⁴ The splanchnic nerves, on the contrary, appear to be without effect upon pancreatic secretion initiated by chemical stimulation of the intestine. Peptone stimulation causes the secretion of a juice resembling that resulting from vagal stimulation, being of higher specific gravity and with a nitrogen concentration many times greater than that caused by acid, water or secretin. Peptone stimulation is also accompanied by depletion of the zymogen granules of the acinar cells (Ramsay, Thomas and Crider). It is possible that certain other products of digestion may enter the circulation and be carried directly to the gland (humoral mechanism, p. 512, footnote). Bile in the intestine appears to be without effect upon pancreatic secretion or is inhibitory, the secretion induced by soap or peptone is definitely inhibited by bile.

The adaptation of the concentrations of amylase and trypsin in the pancreatic juice and tissue to the predominant element of the diet

Grossman, Greengard and Ivy found that the amylase concentration in the pancreatic juice and tissue of rats maintained upon a diet predominantly carbohydrate was greatly increased and the trypsin content decreased. In animals on a high protein diet, the trypsin content of the pancreas was increased and the amylase

⁴ It may be, as Greengard and his associates suggest, that the vagus exerts its effect simply through causing vasodilatation.

content reduced. A high fat diet had little effect, however, upon the production of pancreatic lipase or of trypsin but reduced the concentration of amylase in the pancreatic juice and tissue. The hydrolytic product of starch, namely, glucose, increases the amylase content of the pancreas, but the products of protein (casein) digestion have no such action upon trypsin production. Grossman and his associates suggest that amylase production is regulated through a humoral mechanism involving glucose whereas the tryptic content of the pancreas is controlled reflexly.

THE INTESTINAL JUICE

THE STRUCTURE OF THE INTESTINAL GLANDS

The intestinal juice or *succus entericus* is the name given to a specific secretion of innumerable glands scattered diffusely over the mucosa of the small intestine. The glands consist of minute tubular pits—the *crypts of Lieberkühn*—lined by cells which are continuous with the general epithelial covering of the mucosa. Glands of a similar character are present in the large intestine, including the appendix but their secretion possesses little or no digestive action. Confined to the duodenum, and especially numerous in its upper part, another

type of gland is found, this is the gland of Brunner. The glands of Brunner resemble the gastric glands of the pyloric region and, like the latter, are tortuous and branching. Their secretion, also, like that of the pyloric glands, is alkaline, it contains mucus and a weak proteolytic ferment.

To the naked eye the mucosa of the small intestine presents a velvety appearance, due to the presence of immense numbers of minute slender processes which—like the pile of velvet—project from its surface. These are the *intestinal villi*. They are barely visible individually to the unaided sight and number from 20 to 40 to the square millimeter. They are absent from the large intestine. Each villus shows a lymph vessel (lacteal) which occupies its axis from the base to near its tip. An arteriole enters each villus, and from this a capillary plexus arises from which the blood is returned by one or two venules. The deeper regions of the pits formed between adjacent villi dip into the submucosa and constitute the intestinal glands or crypts of Lieberkühn (fig 401). The cells covering the somewhat club-shaped extremities of the villi are columnar in shape and peculiar in that their free borders show a very fine striation, the *striae* running perpendicular to the free surface of the cell. Many of the columnar cells become

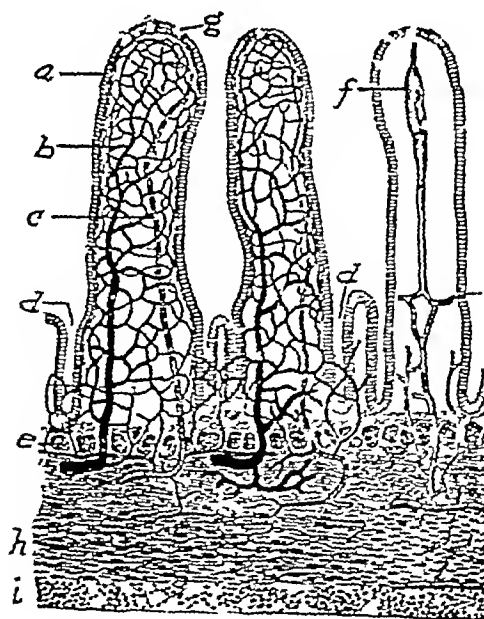


FIG 401 Vertical section through the small intestine showing the villi and intestinal glands *a*, columnar epithelium *b*, vein of villus, *c*, arteriole of villus, *d*, Crypt of Lieberkühn *e*, muscularis mucosae, *f*, lacteal, *g*, goblet cells, *h*, circular muscle, *i*, longitudinal coat (From Bailey's Histology, 13th edition, Williams and Wilkins Co, Baltimore)

transformed into goblet cells which discharge mucus into the intestinal lumen

The cells lining the crypts of Lieberkühn are of four types (a) *Columnar cells* (b) *Goblet cells* These two types of cells are similar to those of the villi but do not possess the fine striations just referred to The more superficial portion of the gland tubule is lined by these two cell types alone (c) Large flask shaped cells which stain with silver ammonium oxide They are therefore known as *argentaffin cells* or *cells of Kultschitzky* These cells contain granules but their function is a disputed question According to some they are endocrine elements, while others believe they supply ferments to the intestinal juice (d) *Cells of Paneth* These lie in the blind extremity of the glands of the ileum but not in those of the duodenum Their protoplasm shows a fine reticulum and is studded with coarse granules The function of these cells has not been definitely determined, but from their resemblance to the secretory cells of other digestive glands they are presumed to supply the ferments of the intestinal juice (fig 40 2)

THE COMPOSITION OF THE INTESTINAL JUICE

Owing to the admixture of the succus entericus with the other intestinal secretions—pancreatic juice and bile—and to the fact that the glands are diffusely scattered throughout the mucosa, special means must be devised for its collection in a pure state Pure juice may be obtained by the establishment of a *Thury fistula* This consists in completely isolating a loop of small bowel from the rest of the intestinal tract by two circular incisions placed from 6 to 10 inches apart, and then restoring the continuity of the bowel by an end to end anastomosis of the sections above and below the isolated portion The blood vessels and nerves reaching the loop through its mesentery are left intact One end of the loop may be closed by suture, the other end being left open and brought through the abdominal wound where it is fastened by sutures and the abdomen closed around it It may be desirable to leave both ends open and stitch each into the abdominal wall a short distance apart This constitutes a *Vella (or Thury Vella) fistula*

The *Mann-Bollman fistula*, so called after its inventors, is useful for certain types of investigation A loop of ileum is isolated The distal (aboral) end is anastomosed to the duodenum or any desired part of the small intestine, the proximal (oral) end is sutured to the abdominal wall By giving the loop this direction in relation to the exterior, the peristaltic waves, which are inherently incapable of passing in any other than an aboral direction along the bowel (p 578) whatever the latter's position, prevent leakage from the external opening

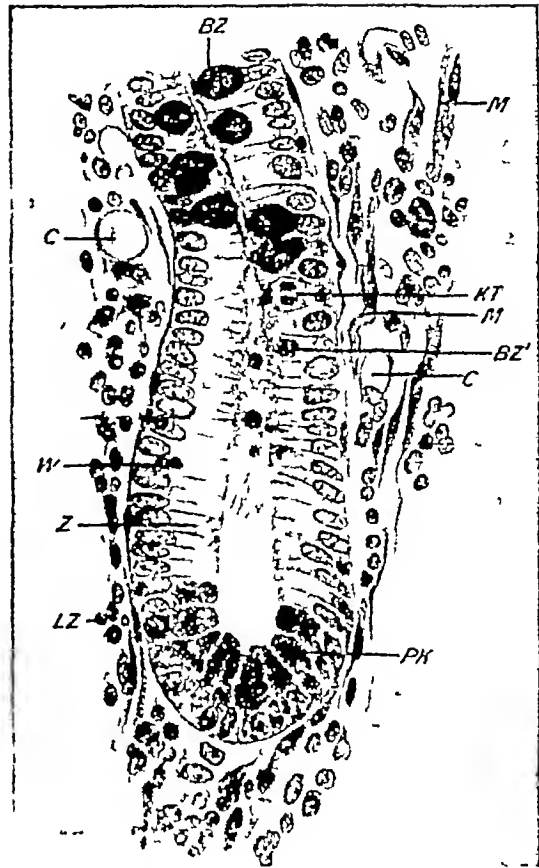


FIG 40 2 A crypt of Lieberkühn with surrounding lamina propria BZ, goblet cells, BZ', goblet cells at the end or beginning of secretion, C, capillary, KT, mitosis in an epithelial cell, LZ, lymphocyte, M, smooth muscle cells, PK, Paneth cells, W, wandering cells in epithelium, Z, epithelial cells of the gland (From Maximow and Bloom, after Schaffer)

In man, samples of intestinal contents can be obtained by means of the Miller-Abbot tube (p 596)

The composition of intestinal juice varies greatly under different conditions At times it consists chiefly of water and salts and has a low digestive power At other times it is slimy from the presence of a large proportion of mucus Its concentration in ferments varies considerably It owes its alkaline reaction (pH from 7.0 to 8.5) to the presence of *sodium carbonate* and *bicarbonate* It contains *enterokinase*, the activator of trypsin (p 527)

The chief ferments of intestinal juice are — (1) *Various peptidases (erepsin)*, (2) A ferment for each of the disaccharids, *sucrose* (cane sugar) *mallose* (malt sugar) and *lactose* (milk sugar), (3) *Lipase*, (4) *Amylase* in traces

Peptidases, erepsin

The peptides that have resisted pancreatic digestion are finally broken up into the separate amino-

acids by peptidases in the intestinal juice. The juice contains several peptid-splitting enzymes, each with a specific action upon a particular type of peptid. The *succus entericus* is almost powerless to attack native protein or peptone, the protein must first have been partially digested by trypsin and have reached the peptid stage. The peptid-splitting action of the intestinal juice has until recent years been attributed to a single enzyme to which the name *erepsin* was applied by Cohnheim. The intestinal mucosa itself is rich in peptidases,⁵ their presence here would appear to afford a safeguard against the entrance into the blood stream of incompletely split products of protein digestion. Peptidases are also found widely distributed, though to variable extent, throughout the tissues of the body, as well as in various plants, yeast and even in bacteria. They are also present in low concentration in pancreatic juice. The optimum pH for the reaction of the peptidases of the *succus entericus* is around 8.0.

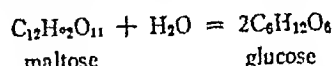
The intestinal wall also contains such enzymes as *nucleases*, *nucleotidases* and *nucleosidases* (ch. 48). The first of these hydrolyze nucleic acids into their constituent nucleotides, the nucleotidases split the nucleotides into nucleosides and phosphoric acid, the nucleosidases separate the nucleosides into sugar (pentose) and purine bases. *Arginase*, a ferment which splits arginine into ornithine and urea (p. 633) is also present in the intestinal mucosa as well as in the liver and other tissues. *Phosphatase* (p. 866) is found in relatively high concentration in the mucosa of the small intestine.

Ferments which act upon sugars

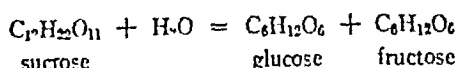
If we except the small amounts of maltase contained in saliva and pancreatic juice, and the comparatively slight change that the HCl of the gastric juice exerts upon cane sugar, then the responsibility for the digestion of the disaccharides devolves entirely upon ferments in the *succus entericus*. *Maltase*, *sucrase* (also called *invertase*) and *lactase* are powerful enzymes which act specifically upon the corresponding substrates to convert

⁵ Though it is generally stated that these and the other enzymes in fistula juice are *secreted* it must be admitted that the experimental conditions under which the juice is collected do not exclude the possibility that its activity is due to enzymes liberated from disintegrated cells of the intestinal mucosa. Some indeed hold the view that under physiological conditions only enterokinase and amylase are actually secreted in the *succus entericus*.

them into monosaccharides. A molecule of maltose is converted by maltase into two molecules of the hexose, glucose, by a process of hydrolysis according to the equation—



Similarly sucrose is split into a molecule each of glucose and fructose by *sucrase*



Lactase converts lactose (milk-sugar) into the hexoses, glucose and galactose. The enzymes acting upon the disaccharides are also present in the tissue of the intestinal wall.

Lipase

The intestinal juice contains small but not unimportant quantities of lipase, relatively large amounts of fats are digested in the absence of the pancreas or after ligation of its ducts. Fifty per cent or more of the ingested fat may undergo hydrolysis under these circumstances.

The various digestive ferments are secreted in much smaller amounts in the lower reaches of the small intestine, and in the large bowel the secretions are composed mainly of water, mucus and inorganic salts, ferments are almost entirely lacking.

THE SECRETION OF INTESTINAL JUICE

The intestinal juice is secreted continuously though in variable quantities from time to time. The activity of the intestinal glands is under *nervous* and *hormonal* influences.

Nervous influences. *Mechanical stimulation.* The glands respond very promptly and energetically to mechanical stimulation of the intestinal mucosa. The secretion is then brought about reflexly through the nerve plexus within the bowel wall. Juice drains continuously from a loop of bowel opening to the exterior and division of the extrinsic nerves does not reduce the secretion resulting from mechanical stimulation (e.g., distention). The secretion indeed is more likely to be increased by sympathetic denervation of the loop, an observation which suggests the removal of an inhibitory influence. Belief in a sympathetic inhibitory influence is supported by the well known experiment of Moreau. He obstructed the bowel at four points by means of ligatures, thus forming three closed loops. The central loop was denervated. The in

testine was replaced within the abdomen which was then closed. Upon examining the loops in from four to sixteen hours later, the central loop was found distended with intestinal juice while those on either side were comparatively empty. It is possible, however, to explain this result upon a vascular basis—namely, the loss of vasomotor tone and consequent hyperemia of the denervated loop—without assuming that an inhibitory influence upon secretion had been removed.

Experimental stimulation of the vagus nerve has on the whole given inconclusive results. Wright and his associates have reported that vagal stimulation causes secretion from the upper part of the duodenum but from no other part of the intestine if the sympathetics were intact. After section of the sympathetics in the thorax a secretion from the intestine followed stimulation of the vagus.

As a result of the sensitivity of the mucosa to mechanical stimuli, undigested food residues or other unabsorbable solid material will in their passage through the bowel cause excitation of the glands and so increase the fluidity of the intestinal contents. The advantage of this in aiding the movement of food along the alimentary canal is obvious.

Hormonal influences It is probable, though it has not been definitely proved, that secretin serves as a stimulant for the secretion of intestinal juice as well as for pancreatic secretion, according to Florey and Harding it excites the glands of Brunner. Injection of secretin into the blood stream has been reported to cause secretion from a Thiry fistula, the secretory effect is unaltered by denervation of the loop of bowel. Secretion into an isolated loop has also been observed when acid was placed in a second loop. The secretion was due presumably to the passage of secretin into the general blood stream. Pancreatic juice acting locally within the bowel lumen, according to Pavlov, also excites the intestinal glands. The juice which results is rich in enterokinase. Pancreatic juice injected into the blood stream has no such effect. Nasset and associates have isolated a substance from the mucosa of the small intestine which they believe is a specific hormone controlling the secretions of the intestinal glands. This substance, which they have named *enterocrinin*, is free from vasodilator effects and does not stimulate pancreatic secretion. It is prepared from extracts of the small and large intestine, it increases both the volume and enzyme concentration of the succus entericus. But un-

equivocal evidence that enterocrinin is a hormone—a principle liberated into the circulation under physiological conditions—has not been secured.

Earlier experiments with secretin seemed to show that it stimulated Brunner's glands, for when injected into the circulation, these glands secreted, and food or acid placed in the stomach or intestine caused secretion from an isolated denervated pouch of the first part of the duodenum (see Ponomarev, Florey and Harding). The effect upon Brunner's glands is due, however, to another substance, termed *duocrinin* by Grossman, for it does not occur with highly purified preparations of secretin. Experiments fall short of establishing duocrinin as an intestinal secretory hormone since the effect on an isolated bowel loop may be due to the absorption of secretagogues.

Secretion of mucus Mucus is secreted by the epithelial cells (goblet cells) of the small intestine and colon. Florey in experiments upon cats concluded that the secretion of mucus by the colon was not under nervous control. Prolonged stimulation of parasympathetic fibers (pelvic nerve) or of the sympathetic was without effect. Pilocarpine, a parasympathetic drug, is also ineffective unless administered in relatively enormous doses. Mechanical stimulation of the mucosa or the application of chemical irritants, e.g., silver nitrate, mustard, iodine, or high temperature, causes a pronounced secretion of mucus. The secretion is brought about, it is believed, by direct stimulation of the goblet cells which increase in number progressively from the upper to the lower regions of the intestinal tract. The injection of histamine, acetylcholine, adenosine or peptone solutions which cause vasodilatation did not cause secretion. The chief functions of the mucus secreted by the colon are to protect the mucosa from injurious agents and, under certain circumstances, by lubricating the intestinal lining, to aid the movement of feces.

A great increase in the production of mucus by the colon is an outstanding feature of the condition known as *mucous colitis*. Abdominal pain, spastic constipation intermitting with attacks of diarrhea and the passage of masses of mucus characterize the disease. Its cause is obscure.

Anthelones (*L. anti* + *G. helkos*, ulcer) This is a principle in extracts of intestinal mucosa which prevents the development of ulcers in dogs after a Mann-Williamson operation (p. 520). A similar substance is present in urine. The intestinal principle is called *enteroanthelone*, and the urinary principle *uranthelone*.

TABLE 38

Water	976.22
Solids	23.78
Mucin and pigments	5.00
Bile salts	9.00
Fatty acids from soaps	1.23
Cholesterol	0.63
Lecithin	0.60
Fat	
Inorganic salts	7.32

THE BILE

In most animals and in man, bile is secreted continuously by the liver cells. It passes along the bile capillaries, hepatic and cystic ducts to be stored in the gall bladder. Its expulsion from the latter and its passage along the common duct into the duodenum is intermittent, related in time to the arrival of food in the intestine, and is quite independent of the actual secretion by the liver (see p. 553).

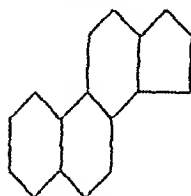
THE COMPOSITION OF THE BILE

The bile is a highly complex fluid. The physiological significance of many of its constituents is unknown, some are present only in minute amounts and are most probably merely waste products undergoing elimination. Table 38 gives the composition of human liver bile (parts in 1000), modified from Hammarsten. Liver bile has a pH of between 8.0 and 8.6. The reaction of human gall-bladder bile is neutral or slightly alkaline, that of the dog (or cat) is definitely acid—pH between 5.0 and 6.0. The chief biliary components are the bile salts, bile pigments, cholesterol and lecithin. These organic materials make up over 60 per cent of the total biliary solids. As a result of the absorption of water and inorganic salts (p. 551) gall-bladder bile is several times more concentrated in organic solids than liver bile. The biliary constituents may vary independently of one another.

The bile salts

The bile salts are the glycocholate and taurocholate of sodium. The bile acids glycocholic ($C_{26}H_{42}NO_6$) and taurocholic ($C_{26}H_{46}NSO_7$), with which the base is combined, are formed by the junction of glycocholl (glycine) and taurin ($C_2H_7NSO_2$), respectively, with cholic acid ($C_{26}H_{46}O_6$). The glycocholic and taurocholic acids are present in about equal amounts in the bile of the herbivora and of man. Taurocholic is the

sole acid in the bile of the dog and other carnivora. Taurin is related to cysteine, a sulphur-containing amino-acid, glycine we are familiar with as the simplest amino-acid and one which can be synthesized by the body (p. 631). Glycocholic and taurocholic acids therefore contain nitrogen, but sulphur is present in the latter only. The structural formula of cholic acid contains the tetracyclic carbon group characteristic of the sterols, this is shown below in skeleton form.



Cholic acid is therefore related to cholesterol, to the male and female sex hormones (chapter 61) and to corticosterone and other adrenal steroid fractions (p. 838). It is highly probable that cholesterol is a precursor of cholic acid in normal metabolism. Block and his associates have demonstrated the production of cholic acid from cholesterol. The latter containing deuterium was fed to dogs, cholic acid containing the isotope was isolated from the urine.

Little can be said concerning the site of origin of the cholic acid, whether it is formed by the hepatic epithelium, or is merely brought preformed to the liver from other body tissues is not known. That some is formed in the body is indicated by the fact that the bile salts continue to be discharged from a biliary fistula during long periods of starvation. That it is derived also from the food appears from the observation that increased excretion follows the ingestion of protein material. Though the supplies of glycocholl and taurine within the body are apparently plentiful the supply of cholic acid is limited, for experiments in which taurine was fed alone caused no increase in the excretion of bile salts whereas cholic acid ingestion alone caused a rise in the excretion of taurocholic acid. It is possible to deplete the taurine stores by feeding cholic acid for several days to a dog with a biliary fistula. When cystine disulphoxide, cysteine, cysteine sulphonic acid or cysteic acid is then fed with cholic acid, an increase in the taurocholic acid of the bile follows (Virtue and Doster-Virtue). This result suggests that from such or similar compounds, taurine is produced in the body. That is, taurine under these circumstances is evidently supplied from body sources whereas cholic acid must be furnished in the diet. The quantity of cholic acid available apparently determines the level of bile acid production (Whipple).

So far as is known the liver is the only situation where the *conjugation* of taurine or of glycochol with cholic acid, and the production of the respective bile acid can take place. The following observations suggest that their formation is a specific function of the liver (1) When the function of the liver is depressed by injury, or by the establishment of an Eck fistula (p 336), the output of bile salts may be reduced by 50 per cent or more (Smyth and Whipple) (2) When the common bile duct is ligated in dogs bile acids appear in the blood. On the other hand, no accumulation occurs in the blood after removal of the liver

THE CIRCULATION OF THE BILE SALTS After their passage into the intestine, the bile salts undergo reabsorption and are carried in the portal blood stream back to the liver for re-excretion. Thus portal-biliary circulation of the bile salts is intimately connected with the absorption of fat (p 541). When bile salts are fed to an animal they can be recovered almost quantitatively from a biliary fistula. This indicates that in the intact animal the reabsorption of bile salts is almost complete (about 90 per cent). Under ordinary circumstances comparatively small amounts (about 10 per cent) of bile salts are formed afresh, i.e., their concentration in the bile is maintained largely as a result of their being circulated over and over again through the portal and biliary systems. Nevertheless, if the bile be prevented from entering the intestine by draining it to the exterior through the fistula, its concentration in bile salts does not become materially reduced. Berman and his associates found that, in dogs which lost all their bile through a fistula, about 0.5 gram of cholic acid was excreted in 8 hours. These facts indicate the existence of some unknown mechanism controlling bile salt production.

Test for bile salts Pettenkoffer's test Five cubic centimeters of the fluid to be tested are mixed in a test tube with a few drops of a 10 per cent solution of cane sugar. A cubic centimeter or two of concentrated sulphuric acid are introduced beneath the surface of the mixture. The appearance of a red ring at the junction of the two liquids indicates the presence of bile salts, upon agitation the color diffuses through the solution. The color is due to the formation of cholalic acid from the bile acids and its combination with the furfural resulting from the decomposition of the cane sugar. A few drops of a 1 in 1000 aqueous solution of furfural itself may be employed (Mylin's modification) instead of the sugar solution.

Rowntree and associates have developed a method for the quantitative estimation of bile salts in blood based upon the Pettenkoffer reaction. The bile salts are

extracted from the blood with alcohol and the test performed upon the extract. The color is compared in a colorimeter with that produced under similar conditions in a standard solution of pure glycocholic acid. The results are expressed in terms of glycocholic acid. Normal human blood contains from 2.5 to 6 mg per cent of bile acids. In obstructive jaundice the value is increased, sometimes several fold.

Hay's test is based upon the property of the bile acids to lower surface tension (p 541).

The bile pigments

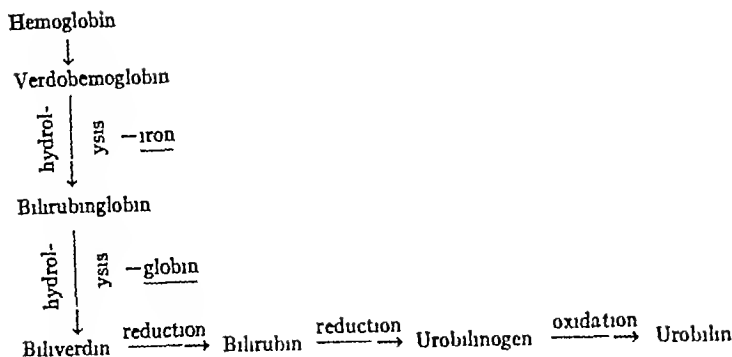
The biliary pigments are *bilirubin* and *biliverdin*. Bilirubin ($C_{33}H_{36}N_4O_6$) is the chief pigment in human bile and in the bile of the carnivora. Biliverdin ($C_{33}H_{36}N_4O_8$) is an oxidative derivative of bilirubin and is present only in small amounts in human bile but is the chief pigment of birds' bile. The pigments constitute from 15 to 20 per cent of the total solids in liver bile.

The readiness with which the bile pigments are oxidized and the color changes which they undergo in the process are the basis for Gmelin's test for bile in body fluids. If, for example, a fluid containing bile be passed through filter paper, and a drop of fuming nitric acid be then dropped upon the wet surface of the filter, the pigment is oxidized and a series of concentric rings of different colors appears—yellow, yellowish-red, violet and blue-green from within outwards. Besides biliverdin other derivatives of bilirubin are found in the body. *Urobilinogen* (cf. below and p 539) is a reduction product of bilirubin, upon oxidation it yields *urobilin*. *Bilicyanin* and *bilifuscin* are formed by the oxidation of biliverdin. The latter two pigments are not found in bile but may be present in gallstones.

THE ORIGIN OF THE BILE PIGMENTS Bilirubin is derived from hemoglobin, being the porphyrin (globin-free and iron-free) fraction of the hemoglobin molecule (ch 6). Any condition which leads to an increased destruction of red cells (hemolysis) causes a greater production of bile pigment.⁶

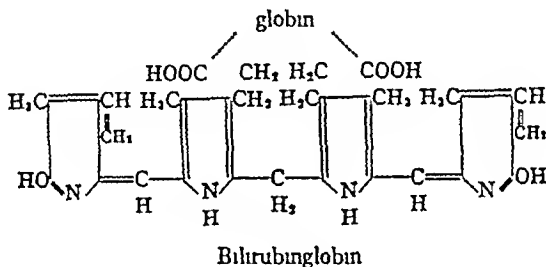
From the researches of Lemberg it appears that an intermediate stage in the production of bilirubin from hemoglobin is the opening by oxidation of the porphyrin ring at the alpha methene bridge. The compound so produced consists of a chain of four pyrrol nuclei and still contains iron and globin (ch 6). It is a greenish pigment called *verde hemoglobin*. Later, the iron is readily split off, the resulting iron-free compound is called *bilirubin-globin*. According to Watson it exists in this form

⁶ Myoglobin (p 58) and possibly pigment constituents of the diet may also be a source of bilirubin.



in the circulation attached to the albumin fraction of the plasma. This complex is responsible most probably for the delayed direct or for the indirect van den Berg reaction. The iron becomes bound to the globulin fraction of the plasma.

was called *hematoidin* (Virchow), the significance of its discovery lay in the fact that the old blood clots in which it was found were in situations remote from the liver. Hepatectomy in geese—which is a relatively simple operation on account of the ana-



The bilirubinglobin next loses its globin—a reaction which occurs in the liver. The pigment now freed of both iron and globin is called *biliverdin*. In man and the carnivora the biliverdin undergoes rapid reduction to *bilirubin*, which is a weak acid and probably combines with sodium to form sodium bilirubinate. The latter is believed to be responsible for the prompt direct van den Berg reaction.

The iron is stored in the liver, spleen and other tissues and with the globin is eventually utilized for hemoglobin synthesis. The chemical relationship between hemoglobin and the various bile pigment derivatives is shown schematically above.

THE SITE OF BILIRUBIN FORMATION has been a subject of controversy for many years and it is only comparatively recently that the question has been answered definitely. The discovery, in old extravasations of blood, of a pigment with the chemical properties of bilirubin, was one of the first observations to throw doubt upon the view then prevailing, namely, that the liver was solely responsible for bile pigment formation. This pigment

tomical peculiarities of the avian liver—was performed by Minkowski and Naunyn with the object of finding a decisive answer to the question of the site of bile pigment production. Arseniuretted hydrogen, a powerful hemolytic poison, when injected into normal birds caused the accumulation of bile pigment in the blood. No such effect followed injection of the poison into hepatectomized birds. This apparently crucial experiment indicated that the liver alone was capable of forming bile pigment. McNee some years later repeated this experiment and obtained the same results, but he suggested that the reticuloendothelial elements of the liver (Kupffer cells) rather than the true secretory cells were responsible. Investigations upon this subject within more recent years have proved conclusively, however, that in mammals the liver is not essential for bile pigment formation. Whipple and Hooper, employing dogs, attempted to isolate the liver from the general circulation. Bilirubin was found in excess in the circulating blood, a result diametrically opposed to the results obtained with birds. The method employed by

these investigators for the isolation of the liver has been criticized, but nevertheless the conclusion which they reached, namely, that tissues other than the liver have the ability to produce bile pigments has been substantiated by Mann, Bollman and Magath, and is now universally accepted. These observers removed the liver completely from dogs (p. 336). The animals survived for 24 hours or more, bilirubin commenced to appear in the blood in from 3 to 6 hours after the operation⁷ and gradually increased in amount up to the time of death. The injection of hemoglobin into the circulation increased the bile pigment accumulation. Mann and his associates also showed by spectroscopic methods that the blood of the splenic vein of a normal animal has a higher bilirubin concentration than that in the corresponding artery.

It is now accepted that the elements of the *reticulo-endothelial system* (see p. 105) situated in various parts of the body, spleen, lymph glands, bone marrow and the general connective tissues, as well as in the liver, are responsible for the formation of bile pigment. Of these, the bone marrow is probably the most important. In an animal deprived of both spleen and liver, bilirubin continues to be formed at approximately the normal rate.

A certain amount of bilirubin is normally present in human serum, (from 0.2 to 0.8 mg. per 100 cc.) as a result of the transformation by the reticulo-endothelial elements of hemoglobin liberated from broken down red cells (p. 73). It has been estimated that 1 gram of hemoglobin yields 40 mg. of bilirubin. The iron which is freed from the hemoglobin is stored mainly in the liver (p. 76) but also to some extent in the spleen. It has been shown by Rich that the hemoglobin-bilirubin transformation effected by the reticulo-endothelial structures is an intracellular process. When fresh red cells were cultured with the reticulo-endothelial cells the former were destroyed, and in a short time typical crystals of bilirubin, and in some cases biliverdin, were found to have formed within the phagocytes. In some instances iron deposition within the cells and adjacent to the crystals could be detected by means of the Prussian blue reaction.

The results of the earlier experiments of Minkowski and Naunyn are explained in the light of these discoveries. The reticulo-endothelial system of the bird is practically confined to the liver. When the liver was

excised all the cells capable of forming bile pigment were therefore removed.

The rôle of the true secretory cells (polygonal cells) of the liver in the production of bilirubin is not known with certainty. The opening of the porphyrin ring and the formation of verdohemoglobin as well as the removal of iron seems to be accomplished entirely by the reticulo-endothelial cells. It is not known whether the removal of globin is a function of the Kupffer cells and other reticulo-endothelial elements or of the hepatic parenchyma. The latter excretes the pigment probably as a sodium salt (Na bilirubinate).

UROBILINOGEN, UROBILIN, THE CIRCULATION OF BILE PIGMENT. The bilirubin that enters the intestine undergoes reduction by bacteria to form *urobilinogen* (also called *stercobilinogen*). A part of this is excreted in the stools and, by exposure to air, is oxidized to *urobilin* (*stercobilin*). The latter can be detected spectroscopically or by the green fluorescence which it gives with a solution of alcoholic zinc acetate. A certain proportion of the urobilinogen is reabsorbed into the portal circulation, passes to the liver, where it is reconverted in part to bilirubin in the bile (fig. 40.3). It is then almost entirely re-excreted both as bilirubin and as urobilinogen. Any urobilinogen which may escape into the general circulation and be excreted by the kidney becomes oxidized to *urobilin* after the urine has been voided. Normally, however, no urobilin or mere traces (0.5 to 2 mg.) appears in the urine. Only traces are present in normal blood. Bilirubin itself does not appear normally in the urine, so the color of urine is due to neither of these pigments (p. 466). That the urobilinogen normally present in bile merely represents re-excretion of this pigment after its absorption from the intestine was clearly shown by Elman and McMaster. When the entire output of bile was collected through a fistula (i.e., none was allowed to enter the intestine) there was complete disappearance of urobilinogen from the bile after the pigment already present in the intestine had been carried away in the feces. The fistula bile remained free from urobilinogen unless infection of the biliary tracts had occurred, under which circumstance bacterial action in these situations caused its formation. When a part of the bile was allowed to enter the intestine or bile was fed by mouth the derived pigment invariably appeared in the fistula bile.

Also, complete experimental obstruction of the

⁷ Dog's plasma normally does not contain appreciable amounts of bilirubin.

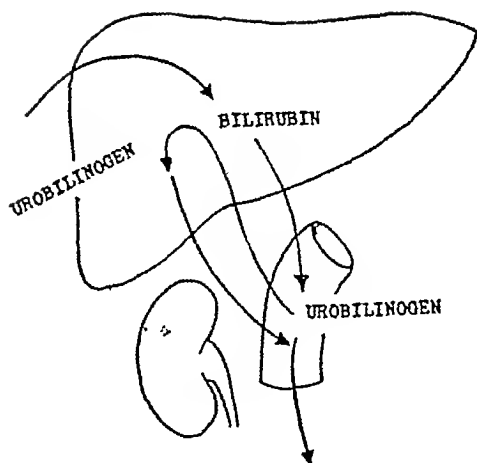


FIG 403 Diagram illustrating the formation of urobilinogen from bilirubin in the intestine, its excretion in part in the feces and its absorption in part into the portal blood. Normally, the absorbed pigment except in negligible amounts is practically all re-excreted in the bile. The dotted lines indicate its passage into the blood, and its excretion by the kidney in cases of liver damage, excessive blood destruction or in the early stages of obstructive jaundice

common bile duct, since it prevented the re-excretion of the urobilinogen absorbed from the intestine, resulted in its accumulation in the blood and its excretion in the urine. These effects, however, can only occur for a short time after the duct has been obstructed, and are due to the absorption of pigment remaining in the intestine from the period prior to obstruction. After the pigment has been cleared from the intestine, the urine, though it may contain large amounts of bilirubin, is quite free from the derived pigment. Depression of the excretory function of the liver by such hepatic poisons as chloroform, carbon tetrachloride, phosphorus, etc., caused urobilinogen to appear in the urine, hepatic damage from other causes, e.g., infectious hepatitis is also associated with the urinary excretion of urobilinogen even though the injury is slight and there is no bilirubinuria. When bilirubin formation is increased by hemolytic agents urobilinogen production is also increased, the liver even though its function is normal is unable to re-excrete the excess pigment absorbed from the intestine, and urobilin appears in the urine.

Urobilinogen is in rare instances formed in situations other than the intestinal tract and independently of bacterial action. Rabinowitch, for example, has reported a case in which large quantities of urobilinogen

were present in the urine. A sterile ovarian cyst containing an old blood clot and a high percentage of urobilinogen was revealed by operation. When the cyst was removed the urobilinogen disappeared from the urine.

Lecithin and cholesterol

Lecithin is present in human liver bile to the extent of from 0.02 to 0.05 per cent. The *cholesterol* content is normally from 0.04 to 0.16 per cent. The cholesterol is present in the free state, i.e., not in the form of esters. The percentages of these materials in the bile of the gall bladder is, as a result of the absorption of water and salts through the gall bladder mucosa, much higher than the percentages in liver bile. The ratio of the concentration of cholesterol in bile to that of the bile salts is from 1:20 to 1:30. The ability of the bile to hold cholesterol in solution is in large degree dependent upon the bile salts. If the ratio falls to 1:13 precipitation of cholesterol occurs. Andrews and associates showed that if bile be dialysed against water, bile salts are removed and the bile becomes turbid as a result of cholesterol precipitation. The importance of this observation in relation to gallstone formation is considered on page 546. The fatty acid concentration of the bile is of as great if not greater importance than the bile salt concentration in holding cholesterol in solution. Little is known concerning the origin and function of the biliary cholesterol. It may be derived from the stroma of disintegrated red cells or from nervous tissue, which of course contains it in large amounts. According to Gardner the cholesterol of the bile does not vary with changes in the total blood cholesterol, nor is its concentration raised by increasing the cholesterol content of the diet; it continues to be excreted upon a cholesterol free diet. A biliary portal circulation of cholesterol has been shown to occur, i.e., excretion in the bile, re-absorption from the intestine and re-excretion by the liver.

THE FUNCTIONS OF THE BILE

Besides serving as a vehicle for the excretion of pigments and certain other waste products from the body, the bile performs important functions in the intestine.

The rôle of the bile in fat digestion

Bile is essential for the efficient digestion of fat by the pancreatic and intestinal juices, yet the bile itself contains practically no lipase or indeed

any other digestive enzyme. The absence of any lipolytic action on the part of the bile is illustrated by the observation of Claude Bernard made upon the lymph vessels of the rabbit's intestine at the height of digestion. In this animal the pancreatic duct opens into the intestine 10 cm. or so below the orifice of the bile duct. After a meal the mesenteric lymph vessels in the intervening area were quite clear, but those below the pancreatic duct were milky with absorbed fat. The importance of bile, nevertheless, in fat digestion (and absorption) was shown by the following experiment. When the orifice of the bile duct in the dog was transplanted into the intestine a short distance *below* the pancreatic duct the lymph vessels of the mesentery between the ducts were free from fat, while those below the transplanted bile duct were milky. The coadjutant action of the bile upon fat digestion is also shown by the fact that, if the bile is prevented from entering the duodenum, either by experimental ligation of the common bile duct or pancreatectomy in animals, or in man as a result of obstruction of the bile flow, by a calculus, tumor of the head of the pancreas, etc., fat digestion is imperfect, a proportion of the fat of the food being excreted unchanged in the feces. The important aid which bile lends to the digestion of fat is due to the bile salts. The influence of the bile salts is exerted in two ways.

First, they aid emulsification to a marked degree, the fat globules are reduced in size, and in consequence, a finer and more stable emulsion results. In this way the total surface area of the fat exposed to the action of pancreatic lipase is enormously increased. This action depends mainly upon the property possessed by the bile acids of lowering surface tension. Upon this property is based Hay's test for the detection of bile in urine. Flowers of sulphur float upon the surface of normal urine but sink if bile is present. The bile salts have also a *hydrotropic action*, i.e., they render water-insoluble materials such as fatty acids and calcium soaps readily soluble in the watery fluids of the intestine. Some of these fatty acids when free in solution aid the emulsification process since they too have the property of lowering surface tension. Their combination with alkali to form soaps—the extent to which this occurs being dependent upon the reaction of the intestinal contents—also, and for the same reason, furthers emulsification. The cholesterol of the bile acts as an adjuvant to the bile salts in the emulsification process.

Secondly, the bile salts (by virtue of the cholic acid radical) act as specific activators of pancreatic lipase, the latter enzyme has only a weak action in their absence. They also act in a similar manner, though to a

less degree, upon gastric lipase, but have no effect apparently, upon the fat-splitting enzyme of the intestinal juice. That this action of bile is distinct from its effect in furthering emulsification is shown by the fact that the hydrolysis by pancreatic lipase of triacetin, which is *soluble* in water, is also accelerated by the bile salts.

The rôle of bile in fat absorption

Fat digestion although seriously impaired may be carried on to a considerable extent in the intestine after removal of the pancreas and ligation of the bile duct, the lipase of the intestinal juice is then chiefly responsible for the cleavage of the fat molecule. Gastric lipase, especially if the intestinal contents are definitely acid in reaction, may also aid to some extent. Certain bacteria in the colon have a lipolytic action and they may account for a considerable quantity of fat being split in the large intestine. In obstruction of the bile ducts alone the greater proportion of the fat is split, over 80 per cent of the fecal fat being in the form of fatty acids. The *absorption* of fat, on the other hand, is very greatly interfered with under these circumstances, only a small proportion of ingested fat being utilized. Bile is therefore of much greater importance for fat absorption than for fat digestion. According to Verzář the fatty acids unite with the bile salts to form complex compounds, as such the greater part of the fat is absorbed from the intestine.

The solvent action of the bile salts upon cholesterol has been mentioned (p. 540). Bile is also necessary for the absorption of vitamin D, the essential constituents for this action being, apparently, the bile salts. Graeves and Schmidt found that the calcium balances of dogs with biliary fistulae became negative but were restored to normal by the subcutaneous injection of vitamin D. Taylor, Weld and Sykes noted that a rise in the serum calcium could not be induced in dogs with biliary fistulae by the oral administration of irradiated ergosterol but that this substance exerted its usual effect when given intravenously, or by mouth if combined with bile or bile salts (see also p. 546). With regard to the other fat-soluble vitamins, vitamin K and probably vitamin E also fails to be absorbed in adequate amounts in the absence of bile. Bile is necessary for the absorption of carotene though not for vitamin A itself.

OTHER FUNCTIONS have been ascribed to the bile, but they are based upon a less sure foundation than those that have been discussed. Among these are (a)

Antiputrefactive This action is probably secondary in nature and not due to any direct inhibitory effect of the bile upon bacterial growth. Indeed it has been shown that bile will serve as a suitable culture medium for microorganisms, and the contents of the gall bladder apparently are readily infected. However, in the absence of the bile putrefactive processes in the intestines are no doubt more active, but this is probably due to the protective covering which the undigested fat affords for protein material, the growth of proteolytic bacteria is thereby encouraged. (b) **Laxative** Bile when introduced directly into the rectum or colon stimulates peristalsis. It is also stimulating to the movements of the small bowel, but the constipating effect which follows the exclusion of bile from the intestine is probably brought about mainly indirectly, the unsplit fat inhibiting intestinal motility. (c) Bile appears to possess a function apart from all those mentioned which is essential to life. Whipple found that bile fistula dogs develop abnormalities of the bones, associated with a loss of inorganic constituents. They eventually succumb unless bile is fed.

THE SECRETION OF BILE

It will be recalled that the secretion of bile by the polygonal cells of the liver and its passage along the biliary channels for storage in the gall-bladder is a continuous process. The intermittent discharge of bile from the gall-bladder will be considered in chapter 41. Several substances stimulate biliary secretion. Materials which act in this way are known as *cholagogues*.^{*} The natural and the most powerful excitants of biliary secretion are the bile salts themselves. Secretin is also a physiological stimulus of biliary secretion (p. 530). Meat or fat in the diet stimulates the flow, whereas sugar is inhibitory. In starvation the volume of the secretion is reduced by about half. Drugs such as podophyllin, aloes, acids and several other chemicals which have enjoyed a reputation as cholagogues, were found by Whipple and his associates to have no notable effect upon biliary secretion. Calomel and ammonium chloride are without any significant effect, but salicylates and linseed oil are definitely stimulating. Adrenaline decreases the flow, whereas pilocarpine and prostigmine increase it. The rate of secretion in man (bile fistula) on an ordinary diet averages around 23 cc. per hour.

^{*} This is a very old term dating from the time of Galen and refers to increased flow of bile into the intestine which, of course, may be due either to increased secretion or to discharge from the gall bladder. It has been suggested recently that the word *cholagogue* be used to indicate a stimulating action upon the discharge of bile from the gall bladder, and the term *choleresis* to denote increased secretion by the liver.

during the waking hours and 15 cc. per hour in sleep—a total volume of around 500 cc. daily.

The bile is capable of being secreted under a pressure of from 250 to 300 mm. of water. When a manometer is placed in the common bile duct, the secretion of bile ceases when the pressure reaches this height and shortly afterward jaundice appears. Under ordinary physiological conditions the sphincter surrounding the opening of the common duct into the duodenum is kept tonically contracted, the duct system fills with bile and when the pressure rises to from 50 to 70 mm. H₂O the fluid flows along the cystic duct into the gall bladder. As a result of the pressure regulating function of the gall bladder (p. 553) the maximal pressure in the biliary canal system is not attained after operation of the common duct in animals until a considerable time has passed. Mann found that jaundice did not develop after this operation in dogs until the lapse of from 48 to 72 hours.

Bile secretion is under vagal control, stimulation of the peripheral end of the vagus (in dog or monkey) causing a very definite increase in bile production (Tantum and Ivy). Snyder found that an acetylcholine like substance appeared in the hepatic veins upon vagal stimulation. A reflex secretory effect can also be demonstrated by stimulating the central cut end of one vagus. Bile secretion is inhibited reflexly by distension of the colon, inhibition also follows stimulation of the sympathetic. This latter effect is probably secondary to constriction of the hepatic vessels. The vagus in the dog also contains some fibers which inhibit secretion. There appears also to be a psychic secretion of bile.

JAUNDICE

DEFINITION AND CLASSIFICATION

When bile pigment is present in excessive amount in the blood (hyperbilirubinemia) it diffuses from the capillaries, the skin, mucous membranes and conjunctivae then become stained a pale yellow tint. But the pigment is not merely dissolved in the tissue fluids but appears to be bound in some way to the tissues. The discoloration is called *jaundice* or *icterus*. The bilirubin appears in the urine and sweat but does not pass into the saliva or milk, nor as a rule into the cerebrospinal fluid.^{*} Jaundice may be due to the production of bile

^{*} Its appearance in the cerebro-spinal fluid is not uncommon in children.

pigment in excess of the amount with which the excretory power of the liver can cope. Or, it may result from the failure of a damaged liver to excrete the bilirubin produced in normal amounts. Jaundice may therefore be divided into two main groups, responding to the mode of its production. There are, the *hemolytic* and the *hepatic*, the former resulting from increased production of bile pigment from hemoglobin, and the latter from decreased or suppressed excretion of pigment by the liver. In the hepatic form the retention of pigment is due to (a) obstruction of the bile passages or (b) damage of the liver cells by toxic agents or infections. Jaundice is also sometimes classified as obstructive and nonobstructive, the latter being subdivided into hemolytic, on the one hand, and toxic or infective on the other. These principal varieties of jaundice are tabulated below.

Hepatic	1 Hemolytic jaundice	} non-obstructive
	2 Toxic, or infective jaundice	
	3 Obstructive jaundice	

A third classification proposed by Rich has come into use based upon whether bile pigment has been taken up by the liver cells and then returned to the blood, or has been rejected by the cells (due to hepatic dysfunction or to an excess of pigment), and thus retained in the circulation. The first mentioned type is called *regurgitation jaundice*, it includes obstructive jaundice, and jaundice due to damage of liver cells, and, therefore of the walls of the biliary canaliculi which permits whole bile to "regurgitate" into the circulation.

In the *second* form, called *retention jaundice*, excessive amounts of bile pigment are produced (hemolytic jaundice) part of which is not taken up by the hepatic parenchyma and is therefore retained in the circulation, or the function of the hepatic cells is so depressed that a large part of the circulating bilirubin, even though not produced in excess, fails to be excreted.

HEMOLYTIC (RETENTION) JAUNDICE

It has already been mentioned that a small amount of bile pigment (0.2 to 0.8 mg per cent) is present in normal human serum and that any condition which increases red cell destruction also increases the formation of bile pigment. However, the functional reserve of the liver is so great that it is very doubtful whether over-production of bile pigment ever taxes the excretory capacity of an

undamaged liver to the limit. But when the hepatic reserve is reduced as a result of disease and hemolysis is excessive the normal balance between the production and the excretion of pigment cannot be maintained, retention occurs and the bilirubin concentration in the blood rises above the normal limits. *Hemolytic agents* of all sorts, such as the toxins of certain infections, septicemia, etc., and various chemical poisons may induce icterus (p. 80) of this type. It also occurs to some degree in such states as *pernicious anemia*, *malaria*, etc., in which blood destruction is a pronounced feature. It may be produced in animals by the injection of such hemolytic poisons as toluidylamine and arseniuretted hydrogen. The disease known as *acholuric jaundice*, which tends to run in families and is associated with splenic enlargement and increased fragility of the red cells (p. 80), is of this type. Jaundice in the new-born, *icterus neonatorum*, (*benign*) frequently occurs as a slight transient staining of the skin and is most probably the result of the destruction of the red cells that are in excess

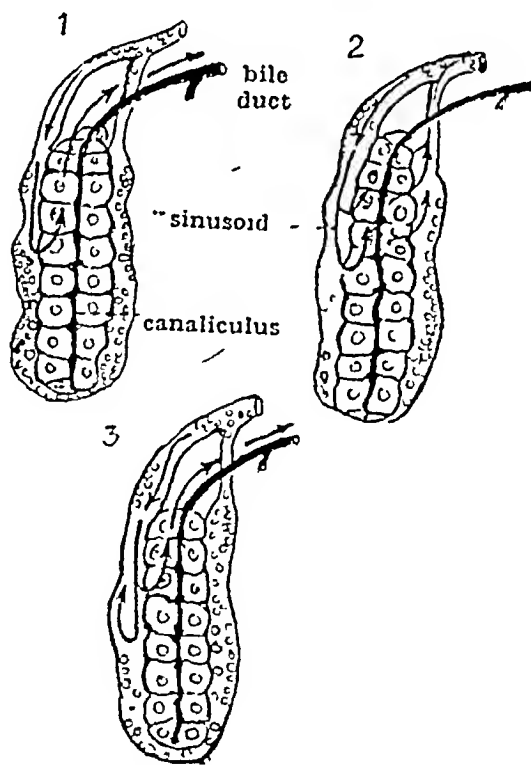


FIG 40.4 Illustrating the course taken by bile pigment, (1) normally, (2) in jaundice due to obstruction or to liver damage (regurgitation jaundice) and (3) in hemolytic (retention) jaundice. In (2) bilirubinate is formed by the liver cells and is absorbed into the circulation, in (3) part of the bilirubin itself or of bilirubinogen is retained in the blood.

at birth (p. 11). It is perfectly innocent and indeed may be considered physiological. It does not appear until a day or two after birth and lasts for five or six days.

OBSTRUCTIVE (REGURGITATION) JAUNDICE

Obstructive jaundice results from blockage of the hepatic or common bile duct by (a) a gallstone or parasites within its lumen (b) compression of the duct by a tumor (e.g. in head of pancreas) or occlusion of its opening into the duodenum (c) congenital obliteration of the ducts (a fatal form of icterus neonatorum).

In complete biliary obstruction the stools are clay-colored, urobilinogen (p. 539) is absent from the feces and urobilin from the urine. In this type of jaundice bile pigment is believed to pass into the liver cells where globin is split off and the acid pigment combined with base to form sodium bilirubinate. The latter after its discharge into the biliary canaliculi is returned—'regurgitated'—into the blood as a result of the rupture or in parenchymatous liver disease disintegration of the canalicular walls.

McMaster and Rous found, for example, that when obstructive jaundice was induced in dogs, the ducts if occluded for long, became filled with a fluid free from pigment and bile salts—the so-called 'white bile' (p. 553). According to these observers the colorless fluid is not bile but a simple sero-mucoid material secreted by the membrane lining the ducts.

The excretory mechanism of the liver has been shown by McMaster and Rous to possess a very large reserve since in the dog jaundice does not develop until from 90 to 95 per cent of the excretory ducts have been occluded.

TOXIC AND INFECTIVE JAUNDICE—LIVER DAMAGE

Liver damage and consequent depression of the secretory functions of the liver may be produced by (a) various poisons e.g., arphenamine, phosphorus, chloroform, etc., (b) acute and chronic liver diseases e.g. infectious hepatitis, homologous serum jaundice (p. 41), acute yellow atrophy, cirrhosis, inflammation of the bile passages (catarrhal jaundice, suppurative cholangitis), (c) toxins of various pathogenic bacteria, (d) engorgement of the hepatic vessels as a result of cardiac failure. It has been pointed out by Meakins that in the latter condition the edema or ascitic fluid does not contain bilirubin and the skin over

edematous regions is not stained. There is no explanation of this fact.

Jaundice is usually due to a combination of causes.

It should be pointed out that in any given case of jaundice two or all three causative factors may and frequently do, coexist. Blood destruction alone, according to Rich, is not capable of producing jaundice except perhaps in rare instances when the hemolytic process is of extreme degree. In pernicious anemia, for example, there is an associated hepatic insufficiency and even in the ordinary icterus of the newborn the immaturity of hepatic function is claimed to be a factor, also, many hemolytic substances are liver poisons as well. In obstructive jaundice the liver cells suffer damage from pressure and the retention of bile salts probably induces a certain amount of hemolysis. Furthermore, in many inflammatory conditions of the liver or bile ducts an obstructive element (due to plugging of the intra-hepatic ducts by the so-called bile-thrombi) exists in conjunction with the hepatic damage. In the jaundice of cardiac diseases the hepatic engorgement produces obstruction of the finer bile capillaries as well as injury to the liver cells through anoxia. The occurrence of infarcts in the lung is also a factor in some cases, bilirubin production is increased through the breakdown of red cells in the infarcted areas.

CLINICAL TESTS EMPLOYED IN THE STUDY OF JAUNDICE

The van den Berg reaction. This test—a modification of Ehrlich's diazo reaction—is employed for the detection of bile pigment in blood serum. There are two main types of the reaction—the *direct* and the *indirect*. The direct reaction occurs without the addition of alcohol, which is essential for the indirect reaction.

The REAGENTS used in the test are

<i>Solution A</i>	
Sulphanilic acid	0.1 gram
Concentrated HCl	1.5 cc
Water up to	100 cc
<i>Solution B</i>	
Sodium nitrite	0.5 gram
Water	100 cc

To perform the test 10 cc. of solution A and 0.5 cc. of solution B are freshly mixed. This mixture is used as the reagent in both forms of the test.

The *direct reaction* is that which follows the addition

of 1 cc. of the reagent to 1 cc. of serum. It may occur in one of three forms, (a) *immediate* or *prompt*, a violet color due to the formation of diazo-bilirubin develops in from 10 to 30 seconds (b) *Delayed reaction*, no change appears until several minutes (5-15 or more) elapsed, then a reddish color develops which gradually deepens to violet (c) *Bi-phasic*, a red color appears promptly as in (a) but takes a variable but longer time than in (b) to change to violet.

The *indirect reaction* is carried out as follows. One cc of serum is mixed with 2 cc. of 95 per cent alcohol. After shaking and centrifuging, to 1 cc. of the supernatant fluid, 0.25 cc. of the reagent mixture and 0.5 cc. of alcohol are added. A reddish-violet color develops almost immediately.

Normal bile and the serum in obstructive jaundice give the prompt direct reaction. Normal serum, the serum in hemolytic jaundice and the bilirubin formed in old blood extravasations into the connective tissues, serous cavities, etc., give the indirect reaction and occasionally a delayed direct reaction. The sera in types of jaundice due to liver damage e.g., hepatitis gives a direct reaction (usually of the delayed or biphasic type) all sera which give the direct reaction also give the indirect but the reverse does not hold. The qualitative van den Berg reaction as a means of distinguishing between the different types of jaundice is regarded today with less favor than formerly.

The indirect reaction used as the basis for the *quantitative* estimation of bilirubin in all types of sera is of much more value. The serum after the color reaction has developed is compared in a colorimeter with a standard solution made by dissolving 2.161 grams of anhydrous cobaltous sulphate in 100 cc. of distilled water. The color of this standard corresponds to that developed by 1 unit of bilirubin. A *unit* is defined by van den Berg as 1 part of bilirubin in 200,000 parts of serum. Normal serum contains from 0.4 to 1.5 unit (i.e., bilirubin is present in a dilution of 1 part in from 500,000 to 100,000 parts of serum, or from 0.2 to 1.0 mg. per 100 cc.). In the quantitative estimation of bilirubin in sera showing the direct reaction the method has been rendered more accurate by the modification of Thannhauser and Anderson. This consists in first adding 0.5 cc. of the reagent to 1 cc. of the serum, and a minute or two later 2.5 cc. of alcohol and 1 cc. of a saturated solution of ammonium sulphate. By adding the diazo-reagent before the alcohol the loss of bilirubin which results from its being carried down with the albuminous precipitate is avoided. When the alcohol is added later the diazo-bilirubin compound is not thrown down but remains in the supernatant fluid.

The quantitative van den Bergh reaction is of value in the detection of latent jaundice, i.e., a hyperbilirubinemia which has not reached the level at which jaundice appears, and in recording the progress of a case of manifest jaundice.

The icteric index The bile pigment concentration may also be estimated by comparing the color of the serum with that of a standard solution. A sample of blood is allowed to clot, after centrifuging, 5 cc. of serum are pipetted off and the color of the sample compared in a colorimeter with a 1 in 10,000 solution of potassium bichromate. The colorimeter is set at 15 mm. for the standard solution. This number is divided by the number on the serum scale when the color of the serum and of the standard solution match. The result is called the *icteric index*. Thus if the reading of the serum scale is 3 the icteric index is 5. The test is invalidated by the presence in the serum of other coloring matter, e.g., carotenoids.

THE FEATURES IN DIFFERENT TYPES OF JAUNDICE COMPARED (SEE TABLE 39)

In obstructive jaundice and in jaundice resulting from liver damage, the staining of the skin, mucous membranes and urine with bilirubin tends to be much more pronounced than in the hemolytic type. Except for a short time after the duct has been obstructed urobilin is absent from the urine and urobilinogen from the feces (p. 539). The plasma alkaline phosphatase shows a pronounced increase in obstructive jaundice, a smaller rise in parenchymatous liver disease, but little or no change in the hemolytic type. In obstructive jaundice the effects, e.g., *bradycardia*, *itching of the skin* (*pruritus*) referable to the retention of biliary constituents other than pigment may be evident. The cardiac slowing has been attributed to the action of the bile salts, though proof is lacking. Since the bile acids are conjugated in the liver it is not to be expected that their concentration in the blood will be raised in a purely hemolytic type of jaundice, this is found to be the case. On this account the latter type is sometimes spoken of as "dissociated jaundice". Also in severe liver damage bile acids may be absent from the blood though the hyperbilirubinemia is pronounced. In other instances of relatively mild degrees of hepatic insufficiency a converse type of "dissociated jaundice", namely, increased concentration of bile acids in the blood without icterus, is sometimes seen.

The bleeding tendency in obstructive jaundice is a serious hazard should a patient be required to undergo a surgical operation. Only within recent years, as a result of the work of Roderick, of Quick and his associates, and of Dam, has the cause of the prolonged coagulation time been discovered. It is due to prothrombin deficiency, which is the

TABLE 39
Chief Differences in the Three Types of Jaundice

	TYPES OF JAUNDICE		
	Parenchymatous liver damage	Hemolytic	Obstructive (complete)
Serum bilirubin	moderate increase	slight or moderate increase	great increase
Urinary bilirubin	moderate increase	little or no change	great increase
Urobilinogen in urine	great increase	moderate increase	absent
Bilirubin and urobilinogen (stercobilin) in feces	moderate increase	great increase	absent
Albumin/globulin ratio	greatly reduced or reversed	little change	little change
Van den Berg Test	delayed direct or biphasic	indirect	direct
Plasma alkaline phosphatase	little change	little change	great increase
Thymol turbidity, cephalin flocculation and galactose tolerance tests	positive	negative	negative
Jaundice	moderate or slight	moderate or slight	pronounced, with pruritus

result in turn of a virtual lack of vitamin K (p 120), for when bile is absent from the intestine the vitamin is not absorbed in adequate amounts. The oral administration of vitamin K with bile salts corrects the hemorrhagic tendency. Possibly an additional though minor factor in the lengthened coagulation time is the retention in the blood of certain organic sulphur-containing compounds (e.g., cysteine and taurine) possessing anticoagulant properties (Carr and Foot). Jaundice due to liver damage is also sometimes associated with a prolonged bleeding time and then appears to be due to the incapacity of the liver to manufacture adequate amounts of prothrombin even though there is no lack of vitamin K in the diet or any failure in its absorption. The cause of the pruritus in obstructive jaundice is unknown. Rowntree and associates found that though a high concentration of bile acids in the blood was frequently associated with pruritus this symptom may occur with a normal bile acid concentration or be absent when the latter is high.

In obstructive jaundice, with complete stoppage of the bile flow, urobilinogen is absent from the urine and feces, but bilirubin excretion in the urine is pronounced, whereas, in hemolytic jaundice the reverse is true, urobilinogen is excreted in the urine and feces in excess, but little bilirubin appears in the urine (see table 39). In parenchymatous liver disease the albumin/globulin ratio is reduced and may be reversed due to an increase

in the beta and gamma globulin fractions, the albumin concentration often being reduced as well. In the other types of jaundice the ratio shows little change from the normal though there may be some reduction in the total protein concentration. The reason for the difference in the van den Berg reactions in the two types of jaundice—the direct in the obstructive and liver injury (regurgitation) types, and the indirect in the hemolytic or retention type—is thought to be that in the former, some physical or chemical change has been brought about in the acid bilirubinglobin after it has entered the hepatic cells, namely the removal of globin and the formation of sodium bilirubinate. Probably the alcohol used for the indirect reaction causes a somewhat similar effect which enables the pigment to react with the reagent. The unchanged pigment in the serum of the hemolytic type is sometimes referred to as *hemobilirubin*, and that of the obstructive type, or of bile itself, as *cholobilirubin*.

Some of the differences between the properties of the two forms of bilirubin and certain observations relevant to the foregoing discussion are given in table 39.

GALLSTONES—CHOLELITHIASIS

Gallstones are composed of constituents of the bile which have been thrown out of solution. Cholesterol is present in greater or less amounts in the commoner varieties of gallstone. Some stones

may be composed almost entirely of this substance. In other types of stone, bile pigment or calcium is an important or the predominant constituent. Gallstones therefore differ considerably in size, color and inner structure according to the materials of which they are composed. They are classified according to their compositions into "pure" cholesterol, cholesterol-pigment-calcium, "pure" bilirubin, bilirubin-calcium and calcium carbonate stones. The latter are very rare in man but not uncommon in cattle. The "pure" cholesterol stone contains from 90 to 98 per cent of cholesterol, the remainder being made up of calcium, bile pigments, protein, etc. The "pure" pigment and the bilirubin-calcium stones, on the other hand, contain varying amounts of cholesterol. The cholesterol-pigment-calcium stone is the commonest variety. Stones of this type are usually multiple and, as a result of pressure of one stone against another, show numerous facets upon their surfaces. They contain about 80 per cent of cholesterol which is deposited in cream-colored layers alternately with darker bilirubin-calcium laminae. The great majority of stones are formed in the gall-bladder, but they may also form in the hepatic duct or even in the smaller ducts within the liver.

The formation of gallstones

The mechanism of gallstone formation is not clearly understood but the following are recognized as being the most important factors to be considered: (a) *Injury, especially of an infective nature* to the gall-bladder wall, (b) *disturbance in cholesterol metabolism*, (c) *stasis* of the bile, and (d) *reaction of the bile*.

(a) **INFECTION** Following the classical work of Naunyn, infection was looked upon as the main, if not the only, cause of gallstone formation. Naunyn maintained, and it is now current teaching, that the cholesterol of the bile was not dependent upon the cholesterol level of the blood and could not be altered by diet. He also claimed, however, that abnormalities of cholesterol metabolism played a minor rôle in the production of gallstones. The mucosa of the gall-bladder, he believed, normally secreted cholesterol and calcium and the secretory process was stimulated by any local inflammatory state. Experiments in which human cholesterol stones were shown to be dissolved after a time in the gall-bladder of the dog under sterile conditions, but not if infection were present, the production of gallstones by injections of microorganisms into the circulation after injury to the gall-bladder, and the frequent occur-

rence of cholelithiasis after infective conditions, notably typhoid fever, were all taken to indicate that infection was essential for the production of biliary calculi. The solvent action of the bile salts upon cholesterol was also held to support this view and to be against the suggestion that this biliary constituent could be thrown out of solution in the absence of infection. It was argued that the bile as it came from the liver could never have so high a concentration of cholesterol that simple deposition could result, but that the cholesterol must be produced in excessive amounts by an inflamed gall-bladder in order to be precipitated. Of the normal concentrating power of the gall-bladder little was then known (see p. 551).

Though not denying the importance of gall-bladder injury and the production of cholesterol from the inflamed mucosa in many cases of cholelithiasis, Aschoff and others have insisted that these conditions are not essential to the formation of calculi, and that certain types, especially the solitary cholesterol stone (see below), can arise in sterile bile and in the absence of any diseased condition of the lining membrane. This view is now generally accepted. Much of the earlier work upon cholesterol metabolism to which Naunyn pinned his faith has been proved to be erroneous.

The multiple cholesterol-pigment-calcium stones are usually looked upon as typical infection stones. They are laminated on cross section and have usually a framework of coagulated protein. These stones are often very numerous, sometimes numbered by hundreds, and an examination of their structure indicates that they have all been formed at about the same time. According to Aschoff, the starting point of their formation is the deposition of pigment in the form of fine rosette-like structures upon which coatings of cholesterol, pigment and calcium are subsequently laid. The inflammatory exudate is rich in protein material derived from the blood, as well as in cholesterol and calcium. The protein, it is pointed out, carries an electric charge of opposite sign to that held by the cholesterol, pigment and inorganic constituents of the bile. It is believed that as a result of these physicochemical relations the deposition of cholesterol combined in varying degree with the other biliary constituents is effected.

(b) **"METABOLIC"** The typical "metabolic" calculus is the large single stone of almost pure cholesterol—the *cholesterol solitaire*. This type of stone, according to Aschoff, is formed quite independently of infection or injury of any sort and is due to the crystallization of cholesterol out of a bile surcharged with this material. The common mixed stones (cholesterol-pigment-calcium) are also composed predominantly of cholesterol and probably in many instances are metabolic rather than infective in origin. In certain conditions, e.g., pregnancy, in which gallstones are prone to develop, the blood cholesterol has been said to be higher than normal (hypercholesterolemia) and there is evidence of a disturbance in cholesterol metabolism. But as a matter

of fact, hypercholesterolemia, according to Gardner, is not common in pregnancy, the normal proportions of free cholesterol to cholesterol esters (cholesterol combined with fatty acids), however, are altered, the former being increased, the latter reduced. Abnormalities in cholesterol metabolism leading to such changes are probably of more importance in the production of gallstones than a rise in the total blood cholesterol. It has already been mentioned that a high blood cholesterol does not cause an increase in the cholesterol of the bile. Also, in certain forms of renal disease and in myxedema in which hypercholesterolemia exists there is little evidence that the latter leads to the production of gallstones, in other forms of renal disease the incidence of gallstones is higher than usual, yet hypercholesterolemia does not occur.

It has been mentioned that increasing the cholesterol of the diet does not raise the cholesterol concentration of the bile, so there is no logical reason, as pointed out by Gardner and associates, for excluding cholesterol rich materials from the diets of those subject to cholelithiasis. Indeed, a high fat diet, by stimulating gall bladder contractions and so preventing undue concentration and stasis of bile, may exert a beneficial effect.

The ratio of cholesterol in bile to the bile salts is an important factor in the formation of gallstones. The cholesterol bile acid ratio in normal bile is between 1:20 and 1:30. Since neither bile salts nor cholesterol are absorbed under normal circumstances from the gall bladder this ratio holds for both liver and gall bladder bile. According to Andrews, deposition of cholesterol occurs when the ratio falls to 1:13. He believes that infection, when it is a factor, plays its part in gallstone formation not so much through increasing cholesterol production as through reducing the bile salt concentration, for he claims that bile salts are absorbed through the inflamed gall bladder mucosa. Dolkart and associates attach more importance to the concentration of fatty acids in the bile than to that of the bile salts in preventing the precipitation of cholesterol.

"Pure" pigment stones (they contain calcium and cholesterol as well) also occur apart from infection. They are small, and dark and, though usually occurring in the gall bladder, may be found in the bile passages. Their origin is not clear, but since they often occur in conditions associated with an abnormally high bilirubin excretion, e.g., acholuric jaundice, they may be the result of the precipitation of bilirubin from a bile which contains excessive amounts of the pigment.

(c) STASIS OR SLOWING OF THE BILIARY FLOW within the bile passages may be responsible for the formation of small stones of pigment-calcium in these situations. When there is complete stasis, due to mechanical obstruction, the fluid in the larger bile passages contains none of the important biliary constituents. The so-called "white bile" fills the ducts and in consequence the formation of calculi is not possible (p. 544). Complete biliary stasis appears also to be a very minor factor in the production of stones in the gall bladder.

(d) REACTION OF THE BILE Until the work of Drury, McMaster and Rous this factor had received comparatively little attention. These observers caused gallstones composed, in varying proportions, of calcium carbonate, pigment and cholesterol to be formed in the bile of dogs which had their gall bladders removed and their common ducts drained into a system of tubing. Encrustations of biliary constituents as well as more or less discrete calculi formed upon the walls of the tubing. These occurred under sterile conditions and in the absence of stasis. The deposits are claimed to result from the alkalinity of the liver bile. Normal bile of the dog as it flows along the bile passages was shown by Okada to be definitely alkaline while that of the gall bladder was acid. Rous and his associates found the liver bile of dogs to have a pH of 8.20 or more, while after its stay in the gall bladder its reaction became decidedly acid—pH 5.18 to 6.00. Bile from the human gall bladder, though less alkaline than liver bile, has rarely a pH below 7.0. One of the functions of the normal gall bladder (p. 553) therefore appears to be depression of the pH of liver bile. So long as this occurs in the usual manner the calcium carbonate of the bile remains in solution. In an alkaline bile, such as is collected from the common duct in the absence of the gall bladder the calcium carbonate is thrown down, and with it the pigment and cholesterol constituents, to be deposited upon the walls of the delivery tubing. These observations suggest therefore that in any condition which interferes with gall bladder function, e.g., infection, injury, or intermittent stasis, the usual acidification of the liver bile will not occur, calcium carbonate will then undergo spontaneous precipitation, and serve as a center or centers upon which the other biliary constituents become deposited.

AN ENUMERATION OF HEPATIC FUNCTIONS

Besides its secretory and excretory functions dealt with in this section the liver plays an important rôle in many other physiological processes. For the reader's convenience a list of these with page references is given below.

- (a) Blood formation in the embryo (p. 103), storage of vitamin B₁₂ (p. 85)
- (b) Fibrinogen production (p. 6)
- (c) Prothrombin production (p. 116)
- (d) Heparin production (p. 115)
- (e) Iron and copper storage (pp. 76, 79)
- (f) Blood volume regulation (p. 335)
- (g) Reticulo-endothelial activity (Kupffer cells) (p. 105)
- (h) Detoxication (p. 592)
- (i) Protein metabolism, deamination (p. 628), amino-acid synthesis (p. 630), urea (p. 632), and uric acid (p. 654), hippuric acid synthesis (p. 458)
- (j) Carbohydrate metabolism (ch. 49)
- (k) Fat metabolism (ch. 50)
- (l) Heat production (p. 736)

(m) Formation of vitamin A from carotene (p 744)

(n) Liberation of a depressor antidiuretic principle (pp 29, 77 and 304)

The hepatic circulation is dealt with in chapter 28

LIVER FUNCTION TESTS

Several of the specific functions of the liver have been utilized as tests for an investigation of the functional capacity of the liver as a whole, i.e., as a means of detecting the presence of hepatic damage and, in some instances, of gauging the extent to which such damage has occurred, and to distinguish between different types of jaundice. Some of these tests will be very briefly described, the accounts being confined in the main to the general principles upon which the tests are based.

A TESTS BASED UPON THE EXCRETORY FUNCTIONS OF THE LIVER

The quantitative van den Bergh reaction has been described (p 544). Since the excretion of bilirubin is an essential hepatic function a determination of the quantity of circulating bilirubin is a valuable means of estimating the extent of liver damage associated with jaundice, provided the hyperbilirubinemia is not the result of biliary obstruction or of blood destruction.

The estimation of the quantity of *urobilinogen* excreted in the urine (p 539) has also been employed as a test of liver function. Increased blood destruction, as in pernicious anemia, quite apart from liver damage will also cause urobilinogenuria, but it is usually much less pronounced than that due to parenchymatous liver disease. Also, infection of the biliary passages may increase the urobilinogen output out of all proportion to the reduction of liver function.

Other tests based upon the excretory function of the liver are the *bromsulphthalein*, *bilirubin* and *rose bengal* tests. In each of these tests the respective material is injected intravenously and the rate of excretion estimated from the quantity retained in the serum after the lapse of a specified time, the concentration of the material in the serum is determined colorimetrically. These three substances are excreted practically entirely by the liver and no significant amounts are taken up by the reticulo-endothelial cells. In normal persons less than 5 per cent of bromsulphthalein (5 mg per kilogram of body weight injected) is retained after 45 minutes, or of bilirubin (1 mg per kilogram of body weight injected) is observed at the end of 4 hours. In the case of rose-bengal (10 cc. of 1 per cent solution injected without regard to subject's weight) 50 per cent or more disappears from the serum within 8 minutes after the injection. These tests, like the van den Berg test, will of course be of no value if obstruction of the bile ducts exists, obviously, whether damaged or not the liver cannot then excrete the injected substances. Colorimetric difficulties also render the dye tests inapplicable in the

presence of hyperbilirubinemia from whatever cause. The bilirubin injection test is employed only in the absence of jaundice, for if the liver's power to excrete the endogenous bilirubin is depressed it is a foregone conclusion that it will show a corresponding incapacity to excrete the injected pigment. In the absence of jaundice, however, the bilirubin injection test is one of the most reliable means of estimating the degree of liver damage. Determination of the *serum alkaline phosphatase* level is one of the most sensitive tests of liver function (p 866).

B TESTS BASED UPON THE METABOLIC FUNCTIONS OF THE LIVER

(a) **THE GALACTOSE AND LEVULOSE (FRUCTOSE) TOLERANCE TESTS** In the *levulose tolerance* test the blood sugar curve is determined after the ingestion of from 40 to 50 grams of pure levulose dissolved in 250 cc of water—the dose of sugar is varied according to the subject's weight. The test is performed in the morning, that is, after a 12 hour fast. The blood sugar level is first determined before the ingestion of the sugar and every half hour for 2 hours thereafter. The liver converts levulose to glycogen, the greater quantity of sugar so converted the less pronounced will be the rise in the blood sugar. In absence of hepatic disease the ingestion of 40 grams or so of the sugar causes a maximum rise of the blood sugar curve of 30 mg per cent or less above the fasting level (when this is between 80 to 100 mg per 100 cc) within one hour, the curve returns to within 10 mg per cent of the fasting level within 2 hours. Definite hepatic injury is indicated by a rise in the blood sugar of over 30 mg per cent when the fasting level is between 80 and 100 mg per cent, and a rise of 35 mg per cent, and of 40 mg per cent when the fasting levels are from 70 to 80 and from 60 to 70 mg per cent, respectively. Failure of the curve to return to within 15 mg per cent of the fasting level after 2 hours, regardless of the height of the curve is also definitely abnormal.

A similar test may be made using levulose. At the end of a 12 hour fast the blood sugar level is determined and 50 gm of levulose given in 400 cc of water. The blood sugar level is then estimated at half hour intervals for 2 hours. Normally the maximum value is reached in one hour and does not rise to more than 30 mg per 100 cc above the fasting level (80–100 mg per cent) and falls to 10 mg per cent above this level by the end of the two hours. The total urinary excretion normally by the end of the 5th hour is less than 130 mg. Higher values than these in blood and urine are found in hepatic insufficiency.

(b) Tests based upon the function of the liver to deaminate the amino-acids with the production of urea (p 632) or upon its detoxicating function have also been devised. Though in the dog hippuric acid is synthesized (from glycine and benzoic acid) only in the kidney in the rabbit and in man this function is performed to an important extent by the liver as well. In carrying

out this test benzoic acid (5.9 grams) is given orally, 1.77 gram of sodium benzoate (equivalent to 1.50 gr benzoic acid) is given intravenously and the hippuric acid excretion determined. Normally, at least 1 gram of hippuric acid is excreted within an hour and the quantity of hippuric acid excreted in the urine at hourly intervals for four hours thereafter is determined. If the liver possesses some reserve from 3 to 3.5 grams are excreted within this time. The determination of the *prothrombin time* (p 119) is also employed as a liver function test. In damage of the liver, the *prothrombin time* is prolonged.

Tests based upon the protein constitution of the plasma

When damage to the hepatic parenchyma exists the albumin/globulin ratio of the plasma is reduced due to an absolute increase in the beta and gamma globulin fractions and usually, as well, to an absolute reduction in the albumin fraction. The tests to be described are based upon the rise in the globulin fractions and the reduction in albumin, for the latter fraction tends to protect the globulin from the action of the reagent and thus to inhibit the reaction. The higher the serum globulin and the lower the albumin the more pronounced will the reaction be. The tests to be described cannot be correlated specifically with any given liver function and are, therefore, largely empirical. The first test of this type to be introduced is known as the *Taka Ara test*, after its originators. In the presence of an excess of globulin a reagent composed of sodium carbonate, mercuric chloride and acid fuchsin when added to the abnormal serum causes the precipitation of the mercuric chloride. A similar test is the *colloidal gold test* of Gray. But the most commonly used and the most satisfactory tests of this group are the *thymol turbidity test* of MacLagan, and the *cephalin-cholesterol flocculation test* of Hanger. In the former, 3 cc of the thymol reagent¹⁰ are added to 0.05 cc. of serum and the degree of turbidity measured after $\frac{1}{2}$ to 1 hour in a spectrophotometer at a wave length of 650 millimicrons. The turbidity is believed

to be due to the precipitation of a globulin-thymol phospholipid complex. The cephalin-cholesterol flocculation test consists of adding an emulsion of cephalin (100 mg from sheep's brain) and cholesterol (300 mg) to the serum and allowing the mixture to stand for from 24 to 48 hours. Normal serum remains clear, whereas the serum from a patient with parenchymatous liver disease (inflammatory, or degenerative) shows flocculation.

Several investigators have studied the reliability of the different liver function tests by comparing the results obtained with the histological findings in biopsy samples of hepatic tissue. Sherlock obtains a cylinder of hepatic tissue by the use of a needle 1 mm in caliber and provided with a trocar. The liver is punctured through the skin under local anesthesia. She finds the most useful and reliable tests were determinations of serum bilirubin, serum alkaline phosphatase and the albumin globulin ratio. In experiments upon dogs in which the liver was damaged by carbon tetrachloride, Drill and Ivy found that the number of hepatic functions affected increased with the severity of the liver damage. The excretion of bromsulphalein was the first to become depressed. Serum alkaline phosphatase showed a rise about the same time or slightly later. Next in order of sensitivity was the test for the *prothrombin time*, the galactose tolerance test was the least sensitive of the four tests employed. From these results Drill and Ivy suggest that the *association* of hepatic functions rather than their *dissociation* (i.e., the singling out of one or other function by liver damage) should receive emphasis.

It is to be remembered that the foregoing are purely functional tests and that a negative result does not necessarily indicate the absence of liver injury. On the contrary, liver disease may exist without its condition being revealed by any of these means. This is obvious from the observations of McMaster and Rous and of Mann and his associates upon the reserve function of the liver. The first mentioned observers showed that in the dog 95 per cent of the excretory function could be abolished before jaundice appeared, Mann and his associates found that the liver tissue of the dog could be reduced by 80 per cent or more without a fall in urea production occurring.

¹⁰ This consists of

1	38 gm	barbitone
1	03 gm.	sodium barbitone
3	gm.	thymol
500	ml	doubly distilled water

CHAPTER 41

THE GALL-BLADDER AND BILE DUCTS

Anatomy

The human gall-bladder has a capacity of about 50 cc. Its wall is composed of a thin layer of muscle fibers and fibro-elastic tissue, with a lining of mucous membrane. The muscle fibers are sparse and loosely interlaced with one another and with the strands of fibro-elastic tissue. The mucosa is surmounted by a layer of columnar epithelium.

The cystic duct through which the bile enters and later leaves the viscus is tortuous, or S-shaped and shows spiral folds of mucosa—the valves of Heister—within its lumen. These so called valves have not a valve-like action, for they offer little resistance to the passage of bile in either direction, nor do they seem to prevent a too rapid flow of bile as some have supposed. They develop late in phylogenetic history being associated apparently with the erect posture, they are found only in primates. The folds are formed in the embryo by the twisting or winding of the duct during development. Their function, as suggested by Keith and supported by Lichtenstein and Ivy, is to stiffen the wall of the duct and prevent its kinking.

The common bile duct, formed by the union of the hepatic and cystic ducts, passes very obliquely through the muscular wall of the duodenum and joins with the pancreatic duct to form the ampulla of Vater (fig 41 1). The latter opens into the duodenum through an orifice situated at the summit of a small papilla about $3\frac{1}{2}$ inches below the pylorus. The ampulla of Vater is surrounded near its outlet into the duodenum by a ring of muscle fibers—the sphincter of Oddi. Boyden has studied the circular fibers surrounding the common bile duct at its duodenal end (i.e., the smooth muscle usually referred to as the sphincter of Oddi) and distinguishes three sets of fibers (a) sphincter choledochus (or Boyden's sphincter), fibers which surround the duct between its penetration of the duodenal wall and the point where it is joined by the pancreatic duct, (b) fibers encircling the pancreatic duct where it opens into the ampulla and (c) those which surround the ampulla itself. These latter are present in only about one-sixth of human subjects, when in spasm they may block the ampullary portion of the common bile duct and permit bile to pass into the pancreatic duct, or pancreatic juice to ascend the common bile duct. The mucosa of the common duct is devoid of the usual mucous glands, but contains special branched, tabular glands lined with tall columnar cells. These glands furnish a thin fluid which dilutes the bile.

The gall-bladder, though it possesses important functions, is not indispensable since it can be removed with

impunity. After such an operation, however, the larger bile ducts undergo dilatation, which may in part compensate for the removal of the viscus. The gall-bladder is absent in some animals whose habits and digestive processes are not essentially different, apparently, from those of animals which possess one. It is absent in the horse, deer and rat but present in cattle, sheep, dogs, cats and mice. It is present in fish, amphibia, reptiles and birds, but is absent from orders lower than these.

CONTRACTIONS OF THE GALL-BLADDER AND COMMON BILE DUCT

The gall-bladder shows spontaneous rhythmical contractions which occur at the rate of from 2 to 6 per minute (in the dog), and also a tonic contraction which lasts for from 5 to 30 minutes or more. The rhythmical contractions (in the dog) are capable of producing a pressure change of from 250 to 300 mm. of water which is about the maximal pressure at which bile can be secreted by the liver (fig 41 2). Rhythmical contractions of the common bile duct have also been demonstrated in animals.

THE FUNCTIONS OF THE GALL-BLADDER

THE CONCENTRATION AND STORAGE OF BILE—THE SECRETION OF MUCUS

Gall-bladder bile may be some ten times more concentrated in total solids than bile collected from the hepatic duct. Water and inorganic salts are absorbed through the lymphatics and blood vessels of the gall-bladder wall. The composition of the absorbed fluid is virtually that of physiological saline. Bile pigments, bile salts and cholesterol are not absorbed to any appreciable degree under normal circumstances.

It is undecided whether cholesterol is excreted by the normal gall-bladder mucosa, though Elman and Taussig present evidence for such a process. In this connection it may be mentioned that a pronounced diffuse deposition of a cholesterol ester in the connective tissue of the human gall-bladder, amounting to from 35 to 60 per cent of its dry weight, is seen as a pathological condition. The tissue of the vesicle is stiff and greatly thickened as a result of its impregnation with lipid material.

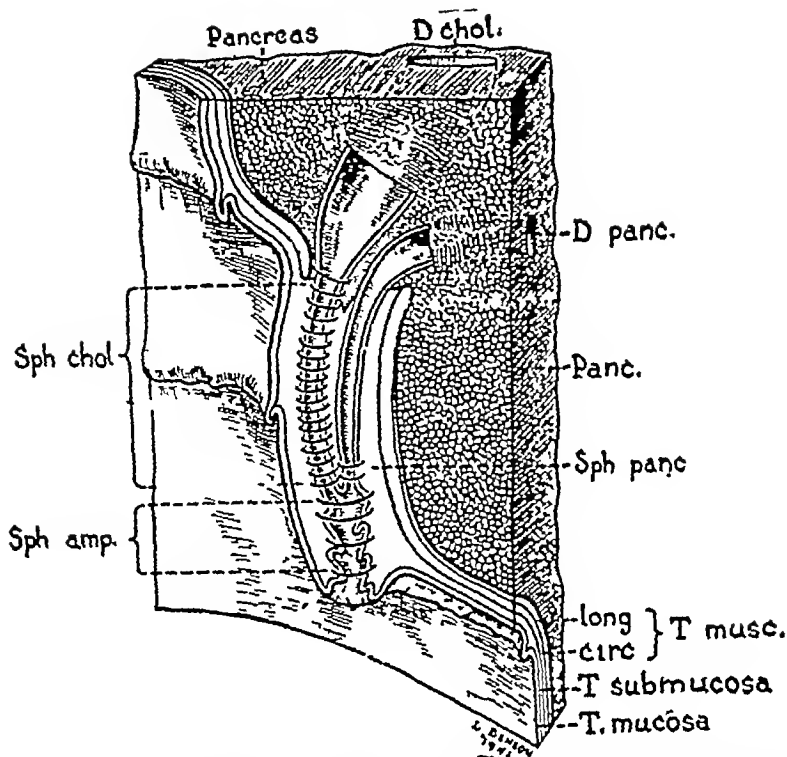


FIG 41.1 Showing the smooth muscle surrounding the duodenal end of the common bile duct. (After Boyden slightly modified.) D chol, common bile duct, D panc, pancreatic duct, Sph chol, sphincter choledochus (Boyden's sphincter), Sph amp, sphincter ampullae, Sph panc, sphincter pancreaticus surrounding the pancreatic duct, T musc., T submucosa and T mucosa, coats of the duodenal wall

The disturbances leading to this condition, which is spoken of as *cholesterosis of the gall-bladder* or, from the appearance of the mucosa, as the "*strawberry gall-bladder*", are unknown, the existence of this condition cannot, however, be used as evidence for the secretion of cholesterol by the gall-bladder mucosa under physiological conditions (See Elman and Graham)

To the work of Hammarsten and more recently to that of Rous and McMaster we owe the greater part of our knowledge of the absorptive powers

of the gall bladder By means of a cannula placed in the bile duct the latter observers collected the bile as it came from the liver and compared its composition with that in the gall-bladder The bilirubin percentages of the respective fluids were used as a measure of the degree of concentration that had been effected in the bile during its stay in the gall-bladder The bladder bile was found to be darker, thicker and more "syrupy" than the bile collected from the ducts It contained from 3.1 to 10.8 times more bilirubin than the liver bile Absorption occurred with remarkable rapidity in some instances In one experiment about 50 cc. of bile which entered the gall-bladder was reduced to less than 5 cc. in about 22 hours In experiments involving the drainage of the gall-bladder through a cannula inserted into its fundus it has been shown that the mere passage of bile through the organ causes a nearly fivefold concentration Inflammation of the gall-bladder reduces or abolishes its concentrating power

The gall-bladder mucosa also adds to the viscosity of the bile by the secretion of a thick

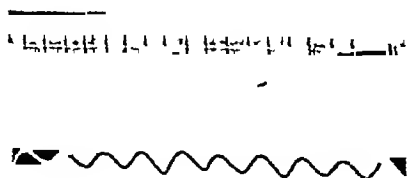


FIG 41.2. Examples of normal gall bladder contractions Time marker, 1 second. (After Taylor and Wilson.)

mucinous material. Little or none of this material is furnished by the bile ducts. Nor were the latter observed to have any concentrating power but were found, on the contrary, to dilute the bile with a thin watery fluid.

When the ducts were obstructed by ligation and the gall-bladder tied off, after some days a clear colorless fluid was found to have collected within the ducts. This fluid—the so-called “white bile”—is not uncommonly seen during an operation upon an obstructed bile duct associated with a functionless gall-bladder. The “white bile” under these circumstances is furnished solely by the mucosa of the ducts. It contains no pigment, bile salts or cholesterol and bears almost no resemblance to bile. The secretion of the latter had been suppressed by the rise in pressure (to 300 mm. or so of water) within the ligated duct system.

If the gall-bladder is healthy and left in communication with the obstructed system the sequence of events is entirely different. Biliary stasis then causes thick, greenish bile to collect in the ducts and bladder as a result of the latter's concentrating activity, and of the mucinous material which it secretes. After a lapse of weeks the imprisoned bile develops an almost tarry consistency. The functions of the ducts and of the bladder are therefore antagonistic, the former tending to dilute, the latter to concentrate the biliary fluid. The diluting effect is at first overbalanced by the concentrating action of the gall-bladder, and when this remains in communication with the duct system, the net result is marked thickening of the bile. There is, however, a tendency with time for the activity of the ducts to overcome that of the gall-bladder. The biliary constituents disappear and ultimately the contents of the system are entirely replaced by the thin simple secretion of the ducts and mucinous material from the bladder (hydrops) (Rous and McMaster).

Another type of “white bile” is sometimes produced. In dogs a clear colorless fluid is secreted by the liver cells when their true secretion is suppressed by some severe liver injury such as that induced by the administration of chloroform (Drury and Rous). “White bile” of this type is sometimes seen in the human subject as a result of hepatic disease.

Other functions of the gall-bladder subsidiary to its concentrating power are the *reduction in the alkalinity* of the bile (p. 548) and the *equalization of pressure* within the biliary duct system. Without the ability to absorb fluid and reduce the bulk of the bile its power to equalize pressure would be negligible. It is to be remembered that the amount of bile secreted in 24 hours is some twenty times

or so greater than could be contained in the gall-bladder. The loss of its action in equalizing the pressure within the duct system is probably a factor leading to the dilatation of the bile ducts, which so frequently follows removal of the gall-bladder (cholecystectomy). After this operation the flow of bile into the intestine is at first nearly continuous, but later the adaptation of the ducts permits intermittent discharge.

The importance of the gall-bladder in the control of pressure within the biliary ducts is apparent from the results of the experiments of Mann and Bollman. They found that after ligation of the common duct in dogs a rise in the bilirubin concentration of the blood did not occur until from 24 to 36 hours had elapsed, and jaundice did not appear for 2 days. If, on the other hand, the gall-bladder was removed at the time that the duct was ligated jaundice was fully developed within 24 hours due, presumably, to the rise in duct pressure, and the “regurgitation” of bile into the blood.

In dogs, cholecystectomy causes some impairment of liver function, as shown by the serum phosphatase test, for at least 70 days following the operation. In man, the excretion of bromsulphthalein is reduced for a short time after this operation, presumably until the ducts can function vicariously for the gall bladder.

THE FILLING AND EVACUATION OF THE GALL-BLADDER

The bile as it leaves the liver flows into the hepatic duct and thence into the common bile duct. During fasting its entrance into the duodenum is blocked by the tonic contraction of the sphincter muscle (sphincter choledochus) at the duodenal end of the duct. As the bile accumulates within the duct its pressure rises, and reaching a height of from 50 to 70 mm. of water, forces its way along the cystic duct into the gall-bladder. During fasting therefore the viscus becomes gradually distended with retained bile.

The nature of the force by which the gall-bladder is evacuated has been a question of some debate. The wall of the gall-bladder is relatively thin, and its muscle fibers so sparse, that it seemed unlikely that it could exert the pressure required to discharge its contents—especially since the gall-bladder is evacuated with considerable difficulty by manual compression. Intra-abdominal pressure, “milking” action exerted by the duodenal move-

ments, and simple leaking into the duodenum as a result of relaxation of the sphincter of Oddi have been variously suggested as possible factors. It has however been proved quite definitely as a result of evidence derived from several modes of investigation that the contractions of the gall bladder itself, despite the apparent muscular weakness of its walls, are responsible for the expulsion of its contents. The times of emptying of the gall bladder are related to gastric digestion. During fasting it remains distended with bile though the sphincter guarding the common duct is relaxed, plainly indicating that the viscus is competent to retain the bile without the aid of the sphincter of Oddi. That changes in intra abdominal pressure are not responsible for its emptying was shown by Mann and Higgins working with guinea-pigs, whose gall-bladder can be readily mobilized.

The abdomen was opened under local anesthesia, the gall bladder exposed and drawn outside the abdominal wound, which was then sutured around the cystic duct. The vesicle was observed to contract and expel its contents in response to food placed in the duodenum. It was also shown that in fish, which of course have no diaphragm and in which, apparently, the intra abdominal pressure remains constant, intermittent evacuations occurred. In dogs the influence of the sphincter was removed by suturing a catheter into the common duct, the abdomen was left open in order to minimize the effects of intra abdominal pressure. The gall bladder remained distended, only during the digestion of a meal did it discharge its contents.

When the walls of the gall bladder contract, bile is discharged along the cystic and common ducts into the duodenum. The sphincter or sphincters guarding the lower end of the common bile duct normally can withstand a pressure of 100 to 120 mm of water but the pressure developed by the contractions of the gall-bladder in dogs was

shown by Mann and his associates to amount to over 250 mm H_2O . It is probable moreover that relaxation of the sphincter occurs as part of a co-ordinated mechanism when the bladder wall contracts, and that the passage of bile through the sphincter is not simply a matter of the latter "giving way" before the biliary pressure created by the gall-bladder contractions. The duodenal muscle surrounding the oblique intramural portion of the common bile duct (p. 551) is capable when contracted of offering a resistance of over 750 mm of water. Since this is much higher than the pressure which contractions of the gall bladder can exert, the flow of bile is completely blocked during contractions of the duodenal muscle but during the latter's relaxation the compression of the duct is relieved (fig. 41.3). Therefore, during the evacuation of the gall bladder and active duodenal movements, the bile may be observed to enter the duodenum in squirts. This is not due to the "milking" action of the peristaltic movements of the bowel but is the result of the alternate blockage and release of the duct, the bowel movements are incapable of causing any flow of bile when the gall-bladder is not contracting.

The most effective stimulus for the discharge of bile is fatty food, particularly egg-yolk, cream or olive oil. It appears that some degree of digestion of the fat must occur before evacuation results. The effect of fat upon the gall bladder was shown definitely by Boyden. He found that during a period of fasting the gall bladder in the cat was distended with bile, and its walls so stretched that they were reduced to about one fifteenth of their thickness in the collapsed state. It emptied slowly after a meal, being collapsed, or nearly so, in from 1½ to 2 hours. The effect of meat upon the discharge of bile is much less than that of fat. Pure protein, such as egg white, and carbohydrate food is almost without effect. These findings have

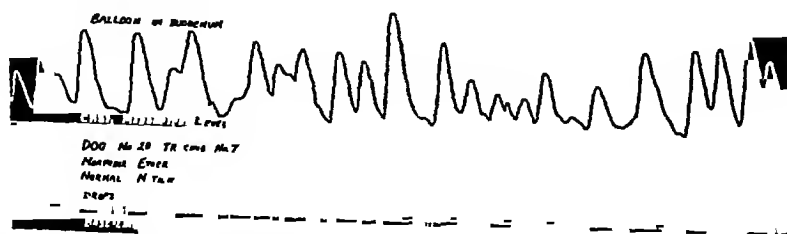


FIG 41.3 This figure shows the relation between duodenal contractions and flow of bile (indicated in drops near bottom of figure) from the common duct into the duodenum (After Lueth)

been amply confirmed by several observers. Whitaker and his associates, for example, observed changes in the contour of the gall-bladder after filling it with iodized oil and examining it radioscopically. The human gall-bladder when rendered opaque to the X-rays by the administration of tetraiodophenolphthalein (p. 556) can also be seen to discharge its contents in response to a meal of fat. Its contractions during operations have also been observed frequently. The products of fat digestion, hydrochloric acid of a strength comparable to that in the chyme, or magnesium sulphate when placed in the duodenum, cause evacuation of the gall-bladder and relaxation of the sphincter of Oddi. Bile salts injected intravenously, on the other hand, cause relaxation of the gall-bladder. Liquid petrolatum introduced into the duodenum is without effect.

The emptying time of the gall-bladder is prolonged in duodenal ulcer, during pregnancy and in pernicious anemia (Boyden).

The mechanisms controlling evacuation of the gall-bladder

(a) **NERVOUS** It has been mentioned that relaxation of the sphincter of Oddi probably occurs as the gall-bladder contracts. A coordinated action of this nature points to a nervous mechanism. The latter may depend upon intrinsic nervous plexuses in the walls of the biliary passages. The gall-bladder contractions initiated by the stimulus of a meal might be due similarly to short reflexes through the intrinsic plexuses of the stomach or duodenum and biliary tract. Nervous mechanisms are also indicated by the following observations. Electrical stimulation of the stomach and duodenum in animals is followed by contractions of the gall-bladder. Contractions are occasionally induced by psychic influences, e.g., the smell or taste of food. The gall-bladder also responds to experimental excitation of the vagus or sympathetic nerves. Experiments attempting to demonstrate the precise actions of the extrinsic nerves upon the gall-bladder movements have, however, given very conflicting evidence. Several observers have obtained weak motor effects from both vagal and sympathetic stimulation, a motor action of the latter is also indicated by the fact that adrenaline is excitatory.

The results of the experiments of Johnson and his associates, however, seem clear cut. Section of the right vagus nerve of the cat retarded empty-

ing of the gall-bladder as a result, apparently, of the interruption of motor fibers to the gall-bladder and of inhibitory fibers to the sphincter of Oddi. The left vagus was found to carry motor fibers to the gall-bladder, but apparently does not contain inhibitory fibers for the sphincter. Reflex effects upon the movements of the gall-bladder may be initiated from other abdominal viscera. Stimulation of the cecum, for example, causes inhibition of the movements. This reflex is abolished after section of the splanchnic nerves or excision of the celiac ganglion. In the cat, Boyden has demonstrated an inhibitory effect upon gall-bladder motility of electrical stimulation of the duodenum by means of tied-in-electrodes. But in a similar human experiment (an electrode being passed into the duodenum through a Rehfuess tube) an inhibitory effect could not be demonstrated. From this and from the fact that the rate of gallbladder emptying is little altered in subjects of double vagotomy, Boyden concludes that in man nervous control of the gall-bladder is of little importance as compared with hormonal control.

(b) **HORMONAL** Even in animals nervous mechanisms are not essential to gall-bladder activity, this is evident from the fact that the reaction to the introduction of fat into the duodenum occurs after all nervous connections between the biliary and gastro intestinal tracts, and between the former and the central nervous system have been severed. That gall-bladder contractions can occur under such circumstances suggests, of course, a hormonal or humoral mechanism. Boyden found that the blood of an animal taken at the height of digestion, when injected into a fasting animal causes the evacuation of bile, blood from a starved animal has no such effect.

Ivy obtained an acid extract from the mucosa of the upper part of the intestine which caused contraction of the gall-bladder when injected intravenously into animals. The injection of acid alone is without effect. Nor will fat or its derivatives excite contractions of the gall-bladder when administered intravenously. Acid and other substances which are excitatory when placed in the duodenum or fed therefore act apparently by causing the production or liberation of a hormone in or from the intestinal mucosa. The active principle is related to secretin but is quite distinct from this hormone, for it does not cause pancreatic secretion, and secretin does not cause gall-bladder contractions. In crossed circulation experiments

the introduction of acid into the duodenum of one animal caused contractions of the gall-bladder of the other. Ivy and Oldberg named this hormone "*cholecystokinin*." As little as 0.2 mg. of the solid material prepared from a potent extract causes definite contractions of the gall-bladder. It is free from histamine and other vasodilator substances. Its effect has been demonstrated upon man. The transfusion of blood from a human subject digesting egg yolk has been found to cause evacuation of the gall bladder of the recipient. No effect was observed with blood from a fasting donor.

The actions of drugs upon the motility of the gall bladder. Adrenaline, pitressin, histamine and mecholyl stimulate the smooth muscle of the gall bladder, whereas morphine, ergotamine and atropine are inhibitory.

CHOLECYSTOGRAPHY

Graham and Cole showed in 1924 that if the chlorine radical of a dye such as tetrachlorophenolphthalein, which is excreted selectively by the liver, be replaced by iodine or bromine, the compound, after concentration in the gall bladder, is opaque to the X rays. Sodium tetraiodophenolphthalein or the corresponding bromine compound (sodium tetrabromophenolphthalein) is given intravenously or by mouth in a special coated capsule. After fasting for a period of about 14 hours a radiogram is taken. The normal gall bladder at this time shows a well defined shadow. The gall bladder is then stimulated to contract by means of a meal containing egg yolk and cream, and a second radiogram taken 5 hours later, when the normal organ should be found practically empty. The depth of the shadow after dye administration depends directly upon the concentrating power of the gall bladder. For this reason a diseased gall bladder may throw only a faint shadow or none at all.

Gallstones, which are relatively transparent to the X rays, especially those of a high cholesterol content, show up against the gall bladder shadow. Gallstones containing more than 0.5 per cent of calcium are visible without the aid of an opaque dye.

AFFECTIONS OF THE GALL-BLADDER AND BILE DUCTS

Among the common diseases of the gall bladder are *inflammation (cholecystitis)*, *gallstones (cholelithiasis)* and *new growths*. The factors involved in the formation of gallstones have been considered (p. 547).

In the absence of inflammation, stones in the gall bladder give rise, as a rule, to no definite symptoms, but in their passage along the ducts severe pain—*biliary colic*—may be experienced as a result of the spasmodic contractions of the gallbladder and consequent distension of the duct walls. Vomiting may occur as a reflex

phenomenon. In the dog pain is produced when the gall bladder is distended by a pressure of 540 mm. of water. This, of course, is a much higher pressure than the contractions of the gall bladder can create. Pain is produced, however, by distending the *ducts* with a pressure of 270 mm. of water—just about the maximal pressure which the gall bladder contractions can produce. The cause of the expulsion of the stone from the gall bladder is not altogether clear. Though contractions of the gall bladder have been observed to cause movements of stones within its cavity, and even to force a stone into the cystic duct, gall bladder contractions do not in the majority of instances offer satisfactory explanation for the expulsion of the stone. According to some, inflammation and distension of the gall bladder are important factors leading to the passage of the stone into the cystic duct.

Biliary dyskinesia

It is now generally recognized that biliary colic may occur in the absence of stone, inflammation or of any anatomical abnormality, such as kinking of the cystic duct, which would hinder the expulsion of bile from gall bladder. In such instances the colic has a functional origin, being due to the gall bladder contracting against a sphincter of Oddi in a state of spasm. Normally, as already mentioned, the sphincter relaxes when the gall bladder contracts. In biliary dyskinesia the nervous mechanisms upon which this reciprocal action depends are, apparently, disordered. It has also been claimed that a sphincter like action may be exerted at the junction of the cystic duct with the gall bladder and that spasm of this ring of muscle during contraction of the gall bladder may give rise to biliary colic. Also, sudden distension with saline of the common bile duct of conscious patients causes pain resembling that of biliary colic. The pain is felt in the upper right quadrant of the abdomen and in the interscapular or the right scapular region. Nausea and vomiting may occur. The pain impulses travel by splanchnic fibers and enter the cord by the posterior nerve roots from the 4th to the 9th thoracic segments, inclusive.

Gall bladder disease frequently gives rise to derangements of other organs, particularly of the stomach. Anacidity or hypoacidity and increased motility of the pyloric part of the stomach are commonly encountered. Ivy and Fishback experimenting with dogs found that mild stimulation of the biliary tract inhibited gastric motility and lowered gastric tone. They suggest that in the human subject stimulation of this nature is conducive to gastric flatulence and belching. A sudden sharp distension of the bile duct caused pylorospasm and vomiting in their animals. According to some authorities dyspeptic symptoms in from 40 to 50 per cent of subjects are due to biliary tract disease. Cardiac irregularities may result through reflexes initiated from the gall bladder or bile ducts.

CHAPTER 42

THE MOVEMENTS OF THE ALIMENTARY CANAL

THE PHYSIOLOGICAL PROPERTIES OF SMOOTH (PLAIN, INVOLUNTARY OR NONSTRIATED) MUSCLE OF THE GASTRO- INTESTINAL TRACT

The chief features of smooth muscle whereby it differs from skeletal muscle are (a) sluggishness of contraction, (b) greater extensibility, (c) the exhibition of sustained contraction or tonus, even when isolated from the central nervous system, (d) the power of rhythmical contraction, (e) the possession of a double autonomic innervation (sympathetic and parasympathetic), (f) greater sensitivity to thermal and chemical influences and to certain types of mechanical stimulation, e.g., stretching, but a lower excitability to electrical stimulation, and (g) longer chronaxie

As in the case of skeletal muscle and of heart muscle (p. 253), the force of the contraction of smooth muscle is dependent, within physiological limits, upon its initial length. Thus distension of the wall of the intestine as by the presence of gas or of fluids secreted as a result of a saline cathartic, causes powerful contractions of the bowel wall.

Tonus of smooth muscle is defined by Evans as the resistance which its substance offers to extension. The degree of tonus (T) may therefore be expressed thus —

$$T = F/L$$

where F is the extending force and L the final length of the muscle subjected to the extending force.

The rhythmical contractions of smooth muscle are superimposed upon the tonus state which varies independently of the rhythmical contractions themselves and may be of high, of medium or of low degree. The processes underlying the production of tonus in smooth muscle are not clearly understood though it is possible that they are not fundamentally different from those responsible for the rhythmical contractions. There appears to be no special structural basis, e.g., different types of fibrils, upon which the property of tonus depends. It is probable that, as in skeletal muscle (p. 961), only a proportion of the fibers are contracting at one time during the tonus state.

That is, that groups of fibers contract in rotation.

The chemical changes accompanying the contraction of smooth muscle are probably similar to those occurring during the contraction of skeletal muscle (ch. 52). Glycogen is broken down and lactic acid produced. In the absence of oxygen, lactic acid accumulates in the muscle contracting in nitrogen, and disappears upon the admission of oxygen.

The tonic contraction of smooth muscle is associated with a negligible expenditure of energy. The tonus mechanism is relatively insusceptible to fatigue, heat production and electrical changes are not detectable, and a rise or a fall in the degree of tonus is not accompanied by a corresponding change in oxygen consumption. Evans found, in fact, that a muscle, when in high tonus, used slightly less oxygen than when relaxed. The tonus of the smooth muscle of the gastro-intestinal tract is dependent mainly upon the intrinsic plexuses, though it is influenced through the extrinsic nerve as well. Tonus changes are also brought about through influences, e.g., changes in hydrogen ion concentration, acting directly upon the muscle fibers.

Postural tone

When a hollow viscus such as the stomach is gradually distended its walls become accommodated automatically to the greater volume, though the tone of the muscle at the new length is altered little from that existing before the distending force was applied. The pressure within the bladder or within the stomach (p. 568), for example, does not increase, or does so very temporarily when the contents of these viscera are increased in volume several fold. The tone of the muscular walls becomes adjusted to the "posture" of the viscus, the adjustment is analogous to the lengthening and shortening reactions of skeletal muscle (p. 965).

The inner mechanism whereby this "postural tone" is brought about is obscure. The great increase in capacity which can occur in a viscus such as the stomach is difficult to explain simply by assuming that the individual muscle fibers are increased in length. It has been suggested, therefore, that the muscle fibers, which are disposed in

layers in the walls of the hollow viscera, slip over one another, the wall thus becoming increased in area but reduced in thickness (Grützner)

MASTICATION

This act comprises movements of the lower jaw, lips, tongue and cheeks. Tearing of the food is effected mainly by the incisor teeth, grinding by the molars. The jaw movements consist of elevation and depression, protrusion and retraction, together with side to side motions, they are all controlled through the inferior maxillary division of the trigeminal nerve, they result in the conversion of the food into a state of fine division and its thorough moistening with saliva. The movements of the tongue and cheeks serve to pass fresh food material between the teeth, and to collect it after treatment by the teeth into a bolus suitable for swallowing.

THE ACT OF SWALLOWING (DEGLUTITION)

It has been customary since the time of Magendie to divide the act of swallowing into three stages.

The *first stage* is under voluntary control. The food which has been transformed into a soft mass by the act of mastication is brought into position upon the dorsum of the tongue, and by the action of the lingual muscles is rolled backwards towards the base of the tongue, where it lies just in front of the isthmus of the fauces. The *mylohyoid* muscle (p. 559) then contracts, pressing the tongue against the hard palate and carrying its base, which is also rotated through the arc of a circle having the hyoid bone as its center, sharply backwards. This movement which is effected with great speed propels the bolus with considerable force into the pharynx where it enters upon the second stage of its journey to the stomach. Other muscles, the *hyoglossi*, the *glossopalatini* and the *styloglossi* acting in conjunction with the mylohyoids assist in drawing the tongue backwards. As a result of the muscular movements, chiefly of the mylo-

hyoids, a pressure of 20 cm. of water is developed in the posterior part of the mouth, pharynx and upper part of the esophagus. A negative pressure, however, exists in the anterior part of the mouth. A negative pressure also normally exists in the closed mouth at other times, which aids in holding the lower jaw in the elevated position.

The *second stage* is brief and is occupied in guiding the food through the pharynx and past the openings leading from it. The muscular movements during this stage are purely reflex in nature. Having once passed through the isthmus of the fauces, the further progress of the food is beyond voluntary control, though sometimes material can be returned to the mouth by a special effort of coughing or "clearing the throat." Upon the food entering the pharynx the constrictor muscles contract, the pressure thus brought to bear upon the food forces it into the esophagus, the *palatopharyngeus* and *stylopharyngeus* muscles at the same time drawing the pharynx upwards over the bolus. There are however, three other possible paths or by-ways along which the food may travel. These are (a) *forward again into the mouth*, (b) *upwards and forwards through the nasopharynx*, and (c) *forwards and downwards into the larynx*. The return of the food into the mouth is prevented by the continued contraction of those muscles which forced it into the pharynx during the first stage. The base of the tongue remains elevated and drawn backwards while the *pharyngopalatini* and *glossopalatini* forming the faucial pillars contract and, approximating the latter structures toward the midline, narrow the opening into the mouth. A negative pressure amounting to 35 cm. H₂O or more is created in the pharynx and esophagus during this time (fig. 42.1) thus aiding in the descent of the bolus.

The opening into the naso-pharynx is closed by the contraction of the *levator veli palatini*, *tensor veli palatini* and *uvulae* muscles which raise the soft palate so that its posterior edge is approximated to the posterior pharyngeal wall. When this movement cannot be accomplished as in post-diphtheritic paralysis, bulbar paralysis, etc., attempts to swallow liquids are followed by their regurgitation into the nasal cavities.

The entrance of food into the larynx is prevented by the contraction of the *thyrohyoid* muscle which raises the organ, bringing its opening under the shelter of the epiglottis and the root of the tongue. The epiglottis itself, however, is not essential for

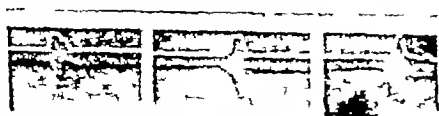


FIG. 42.1 Typical graphs of pressure in pharynx during swallowing. Upper tracing (white), movement of larynx recorded on a tambour. Lower tracing (black), record of pharyngeal pressure. (From Barclay, after Anrep.)

the protection of the laryngeal opening, since when it is destroyed there is little difficulty in preventing the food from entering the air passages—the elevation of the larynx is the important safeguard, when it is fixed by disease swallowing is difficult or impossible. The epiglottis was at one time thought to serve as a sloping lid for the larynx over which the food slid in its passage downwards. This structure, however, stands erect and inclined a little forwards, the food passes over its dorsal (posterior) not over its ventral surface. Coincidentally with the upward movement of the larynx the vocal cords are approximated. This is effected by the contraction of muscles attached to the arytenoid cartilages which are thereby drawn forwards from their usual positions against the posterior pharyngeal wall, and rotated medially. This manoeuvre has the two-fold effect of closing the glottis and causing the upper opening of the esophagus to gape. The food, being subjected to pressure by the contraction of the pharyngeal constrictors, is forced along the path of least resistance.

A short inspiration (inspiration of swallowing) occurs at the very commencement of the first stage, and is followed by complete inhibition of respiration which persists until the end of the second stage.

The *third stage* involves the passage of the food down the esophagus. The food is seized by a peristaltic wave (p. 578) which, travelling along the esophagus, carries the material before it into the stomach. The *cardiac* sphincter guarding the lower end of the esophagus and which at other times is kept tonically closed relaxes upon the approach of the bolus which is then swept into the stomach by the wave of constriction which follows.

Time relationships. The food reaches the upper opening of the esophagus in about 1 second after the initiation of the swallowing act. The rate of progress of the bolus along the human esophagus is not the same at all levels. The muscle in the upper or *cervical portion* of the tube, which is about 6 cm. long is of the striated type and consequently is capable of rapid contraction. Solid food traverses this section in about 1.0 second.¹ In the *upper thoracic* portion, which is approximately 10 cm. in length, the fibers are mixed, some being striated, others unstriated, and the food takes 1.5–2.0 seconds to pass. In its *lower thoracic* part the tube is composed entirely of unstriated muscle. Its length has not been determined precisely but the food takes about 3 seconds to pass through it. Solid or semi solid

food therefore takes between 6 and 7 seconds to pass from the mouth into the stomach.

When liquids are drunk the swallowing process is not quite the same. During the first stage the liquid is squirted forcibly into the pharynx and down the esophagus by the contraction of the mylohyoid and in less than a second reaches the lower end of the gullet which becomes relaxed throughout its entire length. That the fluid is not carried by a peristaltic wave but passes down the esophagus with great rapidity can be demonstrated by means of X-ray or by auscultation over the lower part of the esophagus when the sound of the fluid arriving here can be heard. On this account the accidental drinking of corrosive liquids tends to cause injury, not equally throughout the length of the esophageal mucosa, as would occur if it were carried by a peristaltic wave, but more especially at certain points where the fluid strikes or collects, e.g. in the upper part of the esophagus and above the cardia. When a single mouthful of liquid is drunk the cardiac sphincter does not relax immediately upon the arrival of the fluid. This collects above the sphincter until a peristaltic wave, started in the upper part of the esophagus, arrives a few seconds later to relax the sphincter and carry the accumulated fluid through it. In other instances, the fluid is not shot down the full length of the tube but only as far as the lower part of the cervical segment or into the thoracic portion. The distance depends upon the degree of relaxation of the muscular walls and the force of the propulsive movement. When a series of mouthfuls of liquid are swallowed in rapid succession, as in drinking a glass of water, the esophagus usually relaxes throughout its length, and the fluid is projected to its lower end or even through a relaxed cardia directly into the stomach.

The first stage for the swallowing of liquids is not the same for all species. In the dog and most other mammals the squirting action is important. On this account these animals can drink in the head down position and force the fluid into the esophagus against gravity. This ability is dependent upon the contraction of the mylohyoid aided by the negative pressure in the pharynx and esophagus, removal of the pharyngeal constrictors does not interfere with it. In birds such as fowl, geese, etc. as was shown by Cannon and Moser, the passage of liquids as well as of solids is dependent upon peristalsis, there is no squirting action. These animals in consequence must raise the head in order to swallow liquids which are thus allowed to trickle by gravity into the esophagus. If the mylohyoid muscle is denervated in the dog, liquids must be swallowed in a similar manner.

THE INNERVATION OF THE SWALLOWING REFLEX

Upon first thought, it might appear that the swallowing act could be initiated when desired, yet it is only the first stage that is under voluntary

¹ The esophagus of the dog is composed of striated muscle throughout.

control If one wishes to swallow when the mouth is free of food or foreign material, a little saliva is passed backwards by the tongue and thus serves as a mechanical stimulus for the initiation of the second and third phases of the act which are purely reflex. If the mouth is kept perfectly free of saliva deglutition becomes impossible. The *efferent fibers* of the reflex are furnished by the *hypoglossal* nerve to the lingual muscles, the third division of the *trigeminal* nerve to the mylohyoid, the pharyngeal branches of the *glossopharyngeal* nerve and the pharyngeal and esophageal branches of the *vagus* to the muscles of the pharynx and esophagus

That the carriage of food along the upper two-thirds or so of the gullet is dependent upon extrinsic nerves, and that the peristaltic contraction is not initiated or controlled by nerve plexuses within the wall of the tube, nor yet the result of a property inherent in the muscle itself, was shown by the following experiment of Mosso. When the esophagus was cut across and a swallowing reflex then elicited by stimulating the pharynx, the peristaltic wave which followed "crossed" the gap and appeared in the lower segment. If the *vagus* fibers going to the esophagus are divided, paralysis with complete relaxation and dilatation of the upper two-thirds of the tube results, no peristaltic wave can be evoked, and food does not enter the stomach. On the other hand, the lower quarter or third, in an animal such as the cat of which this portion of the gullet is composed of unstriped muscle, continues to show active peristalsis and may for a time enter into a state of spasm. The cardiac sphincter (p. 561) becomes tonically contracted. Stimulation of the *vagus* causes, as a rule, a strong contraction of the entire esophagus. In the dog whose entire esophagus is composed of striped muscle, section of the *vagi* at any level above the aortic arch causes paralysis of the lower part of the tube which becomes dilated with retained food (Hwang, Essex and Mann). The tone of the cardia is not increased but, on the contrary, is usually reduced.

The sympathetic causes, according to Knight, contraction of the upper and middle thirds and inhibition of the lower third of the tube.

The foregoing experimental results may be taken as evidence that peristalsis in the upper portions of the esophagus is carried out through a central reflex. In the lower esophagus, which in many animals and in man is composed of unstriped muscle,

the initiation and transmission of the wave, like similar movements of the intestine (p. 578), devolves upon the intrinsic nervous mechanism. But also, like the intestine, this part of the tube can be influenced through the extrinsic nerves.

The *afferent fibers* of the reflex are furnished by those branches of the trigeminal, glossopharyngeal nerves and the pharyngeal and superior laryngeal branches of the *vagus*, innervating the mucosa in the regions about the entrance to the pharynx. Usually a certain fairly well defined area in the region of the fauces, the palate or of the posterior pharyngeal wall—the location depending upon the particular animal species—can be demonstrated which is more sensitive for the purpose of eliciting the reflex than any other. In the dog and cat, for instance, the most sensitive region is the mucosa of the posterior pharyngeal wall, supplied by the glossopharyngeal. In the rabbit it lies in the region of the tonsil and the soft palate, the sensory innervation of the latter being chiefly through the trigeminal. The tonsil is also the most sensitive region in the monkey, though stimulation of other areas, such as the entrance to the larynx or the base of the epiglottis will also elicit the reflex. In man, the most responsive area has not been precisely determined, but the reflex is readily induced by mechanical stimulation of the posterior pharyngeal wall or of the base of the tongue. If the impulses arising from these areas are annulled by cocaineization, swallowing is most difficult or even impossible. The central reflex cannot be initiated from the esophagus itself, a bolus placed in its upper or middle third remains in position until the reflex is evoked by stimulating the pharyngeal mucosa.

Experimental stimulation of the central end of the superior laryngeal nerve will elicit the complete series of swallowing movements. The glossopharyngeal nerve contains afferent fibers which excite the reflex, others which inhibit it. A reflex started by the stimulation, say, of the superior laryngeal may at once be cut short by stimulation of the central end of the glossopharyngeal. The tonic inhibitory effect of the glossopharyngeal may be shown by simply cutting the nerve, when the esophagus passes into a state of tonic contraction which persists for some time. The afferent glossopharyngeal fibers responsible for the inhibitory effect respond to a milder stimulus than do those which excite the reflex. The glossopharyngeal also contains those fibers through which the inhibition of respiration is effected reflexly during the descent of the food through the pharynx. Stimulation of the central end of the nerve causes the

breathing to be immediately checked, either in inspiration or expiration according to the phase during which the stimulus is applied. The inhibitory effect of deglutition upon respiration may be demonstrated upon oneself. If the breath be held until the desire to breathe becomes imperative the distress is momentarily relieved by swallowing.

Deglutition centers

The large number of different nerves involved in the act of swallowing and the smooth manner in which the act, once initiated, is carried to completion, indicate that the cranial nuclei are woven together by association fibers into a coordinated whole, each nuclear part of the mechanism coming into action in accurately timed sequence and inevitably, when the first of the series is excited. The central nervous tissue composing this coordinated mechanism is therefore diffuse and extends throughout the pons and medulla. A more circumscribed area or center apparently presides over this complicated mechanism. Markwald found this area to lie in the neighborhood of the vagus nucleus and above but independent of the respiratory center. Here, it is presumed, the afferent impulses are received and relayed to the motor nuclei. In progressive bulbar paralysis the central mechanisms are very gravely affected as a result of the successive involvement of the several nuclei.

ANTIPERISTALSIS IN THE ESOPHAGUS, HEART BURN, BELCHING

According to Alvarez, reverse wavelets or ripples commencing at the cardia and passing upwards along the esophagus to the pharynx are not an uncommon occurrence in man, but true reverse peristalsis is not seen unless some obstruction exists. The reverse ripples in the esophagus are thought to be responsible for some of the symptoms of dyspepsia, e.g., the deposition of "fur" upon the back of the tongue, bad breath and regurgitation of fluids into the mouth. Evidence for the existence of reverse movements in the esophagus was obtained by Kast who found that lycopodium spores swallowed in a capsule were recoverable from the mouth washings next morning in over 50 per cent of a series of human subjects. The possibility of material from even the lower bowel reaching the mouth is strongly suggested by the fact that lycopodium spores introduced into the colon by enema have been recovered some hours later from washings of the stomach.

"HEART BURN" is ascribed by Alvarez to the

stimulation of the mucosa of the upper part of the esophagus by acid fluid regurgitated from the stomach. Payne and Poulton suggest that a tonic spasm of the esophagus set up by the acid stimulus is responsible for the sensation. Jones and Richardson produced it in normal persons by distension of the lower third of the esophagus. The introduction of acid, cold water or gastric contents into this part of the esophagus also caused the burning sensation. Spasm of the wall of the tube was observed at the level in contact with the material and reversed peristalsis above. It is generally agreed that the sensation does not originate within the stomach itself.

BELCHING The tendency, after a meal, for small amounts of gas to be expelled from the stomach into the esophagus and mouth is experienced by most normal persons. It is brought about, most probably, by reverse waves originating in the cardiac region of the stomach and ascending the esophagus. The repeated belching of gas is, however, abnormal. The gas in these instances is not, as a rule, produced by digestive or fermentative processes in the stomach, but is simply air which has been previously swallowed (aerophagy); it has the composition of atmospheric air. The greater part of the swallowed air does not enter the stomach but is held in the lower part of the esophagus until a sufficient volume has collected to give the subject a certain satisfaction when it is belched. The condition is seen in the nervous type of subject or in one who has some gastric discomfort, he resorts to the air-swallowing trick in an effort to gain relief. The intragastric pressure is not increased above the normal, apparently, in subjects who have the sensation of "gas on the stomach" and the pressure is not lowered after gas has been belched.

THE CARDIA

The muscular ring encircling the lower end of the esophagus is commonly known as the cardiac sphincter, though, as a matter of fact, in man the thickness of the muscle in this situation is scarcely greater than that in the rest of the tube. We have already seen that the cardia is held tonically contracted, but relaxes upon the approach of a peristaltic wave or may even remain relaxed during a series of swallows, as in drinking (p. 559).

Experiments upon animals suggest that the cardia is supplied with both motor and inhibitory fibers from the vagus, stimulation of this nerve

though followed in many cases by relaxation of the sphincter, at other times causes contraction. Though in some animals a tonic contraction of the cardiac sphincter follows section of the vagi, Hwang and his associates found that bilateral vagotomy in the dog in no instance caused increased tone of the sphincter, as would result had they an inhibitory function, but hypotonia was sometimes the result of this operation. However, the character of the response, inhibitory or excitatory, depends, apparently, upon the degree of tone, high or low respectively, exhibited by the sphincter at the moment of stimulation. Species differences may also be responsible for the variability of results. Observations upon cardiospasm (q v) suggest that in the human subject *the chief influence exerted by the vagal fibers is inhibitory*.

The cardia is also innervated by the sympathetic, but there has been uncertainty again as to whether this nerve exerts an inhibitory or an excitatory action. As in the case of the vagus, both inhibitory and motor effects have been reported (Carlson) to result from its stimulation. In man the sympathetic has been thought to have a motor action. The fact mentioned earlier, that in some animals bilateral vagotomy causes spastic contraction of the cardia, would also suggest that the usual action of the sympathetic upon the sphincter is excitatory, were it not for the observations of Hwang and associates that sympathetic ganglionectomy is without any effect upon sphincter tone, whereas vagal section may be followed by hypotonia. Knight in experiments upon cats found that whereas vagal stimulation, in all instances, caused relaxation of the cardia, sympathetic stimulation invariably resulted in contraction.

Thus, as in the case of the vagus, the conflict of experimental results in respect to the sympathetic innervation of the cardia does not permit any clear cut statement to be made. The confusion is likely due to species differences. Observations in cases of cardiospasm (see below) have suggested a motor action of the sympathetic in man but Grimson has thrown doubt upon such a function for he could find no change in sphincter action after double supradiaphragmatic vagotomy.

The cardia relaxes much more readily to pressure applied to its esophageal aspect than to pressure from within the stomach. In animals, a pressure of from 5 to 7 cm. of water on its upper surface is sufficient to cause relaxation, but a pressure of 25 cm. is required to be exerted from the stomach side. Alvarez found, however, that in the human

subject the mere pressure of the stomach contents upon the cardia may cause it to relax. This fact was brought out by placing persons in the head-down position. The tonicity varies in degree in different individuals, and in some, simply bending over forces the cardia, and causes a reflux of fluid into the pharynx or mouth, as a result, no doubt, of the compression of the abdomen, and the consequent elevation of intragastric pressure. It is to be recalled in this regard that the pressure in the esophagus is, like that within the rest of the thorax generally, subatmospheric. In individuals in which the tone of the cardia is lessened the "negative" pressure in the esophagus together with any increase of pressure in the abdomen, will encourage the regurgitation of fluid from the stomach.

The tone of the cardia is inhibited by mild stimulation of the gastric mucosa, and by sensory impulses arising in the mouth and pharynx. Its tone increases as digestion proceeds. The factor responsible for the increased tone does not appear to be the acidity of the gastric contents, the hypertonicity apparently is a part of the general increase in tone that occurs in the fundic portion of the stomach with the progress of digestion.

The tone of the sphincter may be *increased* reflexly by abnormally strong stimulation of the stomach or of more remote regions of the alimentary tract. Afferent impulses arising from a diseased gall bladder and other abdominal organs have been held responsible for abnormally hypertonic states of the cardia.

CARDIOSPASM. Cardiospasm is the term applied to a condition in which the sphincter does not relax properly during deglutition, difficulty in swallowing (dysphagia) results, the subject complaining that the food "sticks in his throat." X-ray examination frequently shows that the lower portion of the esophagus is dilated into a funnel shaped or fusiform structure. The condition is usually one of incoordination between the muscle of the esophageal wall and the sphincter—*achalasia*—rather than one of actual spasm. Hurst showed that in most instances a tube weighted with mercury passed readily through the obstruction, and suggested the term *achalasia* in these cases in preference to cardiospasm. The mode of its production is not known for certain. In some instances it may be reflex in nature and due to the irritation of afferent fibers in the stomach, gall bladder or other abdominal viscus. In two cases reported by Rake, from which tissue was obtained for histological examination, degeneration of Auerbach's plexus (which receive vagal fibers) was found. Degenerative changes in the ganglion cells of this plexus have been described by several investigators. Rake considers the condition

analogous to Hirschsprung's disease (p 591) which is due to an imbalance between sympathetic and parasympathetic influences. These findings suggest that the normal action of the vagus upon this part of the alimentary tract of the human subject is to raise the tone of the esophageal wall and lower that of the cardia. Increased resistance at the cardia together with dilatation of the esophageal wall would therefore naturally result from disease of the vagal endings. Knight has reported the treatment of cardiospasm and achalasia by sympathectomy. But this operation which was designed to remove excitatory impulses is not, as a rule, permanently successful.

In some instances achalasia of the cardia or even true spasm appears to be a manifestation of vitamin B deficiency and benefit has followed the administration of thiamine.

MECHANICAL FACTORS IN THE PHYSIOLOGY OF THE STOMACH

GENERAL CONSIDERATIONS SHAPE, POSITION AND DIVISIONS OF THE STOMACH

The normal position of the empty human stomach is not horizontal, as used to be thought before the development of roentgenography. The greater part of the viscus is vertical or nearly so. In some subjects the stomach is shaped like the letter J, in others like a steer's horn and in others again like a reversed L. The majority of normal stomachs are J-shaped. In this type the pylorus lies at a higher level than the lowest part of the greater curvature and the body of the stomach is nearly vertical. In the steer-horn and reversed L types, on the other hand, the pylorus lies on a level with or below the most dependent part of the greater curvature. The body of the reversed L-shaped stomach is also nearly vertical whereas that of the steer-horn type slants from above downwards and to the right (fig 42 2). The hypotonic or atonic stomach tends to assume the J-shape, the hypertonic stomach is of the steer-horn type. The tone of the reversed L-shaped stomach is intermediate between the two.

The lowest point of the greater curvature reaches, usually, to about the level of the umbilicus, or to a line drawn between the highest points of the iliac crests. But even in quite healthy and apparently normal individuals it is often well below the umbilicus in the standing position. A low position of the greater curvature has been looked upon in the past as pathological and has been believed to be the cause of various dyspeptic symptoms. X-ray examinations, however, have shown

that in a large proportion of perfectly healthy persons the lower limit of the stomach may even be well down in the pelvis. But there is no "drooping" or ptosis of the viscus in the true sense, for the fundus remains in contact with the diaphragm. That is, the stomach is elongated but no downward displacement of the organ as a whole occurs. Most observers now believe that the descent of the greater curvature is an unusual finding which, though sometimes associated with digestive disorders is not, as a rule, a cause of them, and may usually be disregarded. The stomach does not empty itself by gravity, but through the contraction of its muscular wall like any other part of the digestive tube, of which it is merely a dilated segment. It is not to be expected, therefore, that a lower position of a portion of a stomach with healthy muscle would necessarily affect its emptying any more than the evacuation of a loop of intestine would be influenced by the position which it occupied in relation to the rest of the intestinal canal. The pylorus lies in or a few centimeters to the right or left of the mid-line. When the stomach fills, the pylorus moves to the right.

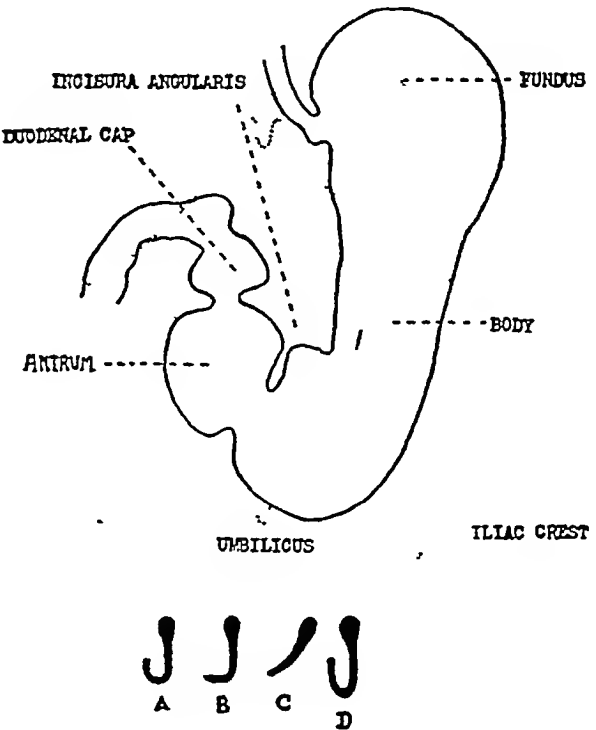


FIG 42 2 Diagram showing the position and subdivisions of the stomach. A, B and C represent J, reversed L and steer horn types of stomach, respectively. D, stomach of the J-shaped type with the greater curvature lying well below the level of the umbilicus. Antrum is also known as the vestibule.

The wall of the stomach is composed of a serous coat and three layers of muscle fibers—longitudinal, circular and oblique, from without inwards. The mucous membrane is thrown into numerous ridges or folds called *rugae*. These are particularly well marked when the stomach is empty but tend to become flattened out as the organ becomes distended. There is a well marked *muscularis mucosae*. The capacity of the stomach is from 1 to 1½ quarts.

The portion of the stomach lying above an imaginary horizontal plane passing through the cardiac orifice is called the *fundus* (fig. 42.2). It is filled with entrapped gas. The more or less vertical portion of the organ below the fundus is called the *body*. The succeeding portion, represented by the hook of the J in the case of the J-shaped stomach, is known as the *pyloric part* (*pars pylorica*) and consists of a proximal chamber the *antrum* or *pyloric vestibule* and the *pyloric canal* through which the stomach communicates with the duodenum. The notch at the lower end of the lesser curvature formed by the bending of the pyloric part upon the body, is known as the *incisura angularis*. The pyloric canal is a narrow passage about 3 cm. long around which the circular coat of the stomach is thickened to form the *pyloric sphincter*. The thickening of the muscle causes a projection of the mucosa which is responsible for the narrowing of the digestive tube at this point and for the slight circular depression on its outer surface. The musculature of the stomach is almost completely separated from that of the duodenum by a ring of connective tissue.

THE MOTOR ACTIVITIES OF THE STOMACH

When the stomach is empty its cavity below the upper part of the fundus, which as mentioned above is inflated with gas, is nearly obliterated by the apposition of the gastric walls. Food, after passing through the cardia, collects just above the



FIG. 42.3 Upper series, X-ray appearance of stomach at 200 second intervals following the ingestion of a liquid meal. Lower series, appearance of stomach at 5, 5, 5 and 10 minute intervals, respectively, after the ingestion of a solid meal. (After Wilson, Dickson and Singleton.)

obliterated portion, apparently simply of its own weight it gradually separates the gastric walls, and passes downwards along the lesser curvature into the body and pyloric part of the organ (fig. 42.3). This passage-way is sometimes referred to as the *magenstrasse*.

The fundus and usually the remainder of the stomach above about the middle of the body shows little or no motility. The muscle of this part of the organ is the seat of a weak tonic contraction which is immediately inhibited by the entrance of food into the stomach or even by the taste of food or its passage down the esophagus (see p. 568). This "*receptive relaxation*" of the gastric wall is known as the *gastric feeding reflex*; it also involves inhibition of peristalsis in the rest of the stomach. The reflex is mediated through the vagus nerves. Welch has observed abnormalities of the reflex in intra-abdominal conditions outside the stomach. Instead of the usual effect of eating upon the activity of the stomach, an increase of tone and of peristalsis occurs, which not infrequently gives rise to minor digestive disturbances, e.g., a feeling of fullness, tightness, eructations, heart burn, or even pain (see fig. 42.4).

The lower part of the body and the antrum of the stomach constitute a chamber wherein the food is macerated, fragmented, and thoroughly mixed. Peristaltic waves commencing near the middle of the body of the stomach sweep downwards through the pyloric vestibule. They are shallow and ill-defined at their commencement but become stronger as they descend. They also increase in strength as digestion proceeds and, when this is at its height, bite deeply into the gastric walls.

Two or more peristaltic waves may be seen at one moment travelling through the lower part of the body and pyloric region, for this reason the X-ray appearance of the actively motile stomach is irregularly convoluted. The waves in their downward journey show rhythmical variations in depth. They might be described as waxing and waning. At one instant they deeply indent the gastric wall and the segments between the annular constrictions are also reduced in diameter. There appears to be a general increase in tone of the gastric wall, the capacity of the entire pyloric region being reduced. At the next instant there is a general reduction in tone, the waves are less intense, the convolutions of the stomach outline less pronounced, and the capacity of the pyloric region increased (fig. 42.5).

Gastric motility shows great individual variation, in some types of stomach the wave travels

very rapidly, completing its journey in from 10 to 15 seconds. In others the wave takes 30 seconds or so to pass from its origin to the pylorus. The slow waves are the more common.

A second type of wave is seen in the antrum which may occur rhythmically or arrhythmically. The former type is of the same rate as the peristaltic waves but much weaker, exerting a pressure of less than 5 cm of water. These waves, which have a duration of about 20 seconds, are not propulsive, being purely mixing or churning in function. Variations in tone also occur in the antrum upon which these contractions are superimposed.

Apart from individual differences, the motor activity of the stomach is influenced by the chemical characters and bulk of the meal. Fat, for example, inhibits the movements, reducing the depth of the peristaltic waves, and lowering gastric tone. Fat exerts this effect after reaching the duodenum. Ivy and his associates have obtained evidence that the inhibitory effect of fat is mediated through a chalone (see enterogastrone, p. 512). Gastric motility is altered in duodenal ulcer, the weak mixing movements being less frequent than normally, while peristalsis becomes more pronounced; they appear to be largely responsible for the ulcer pain.

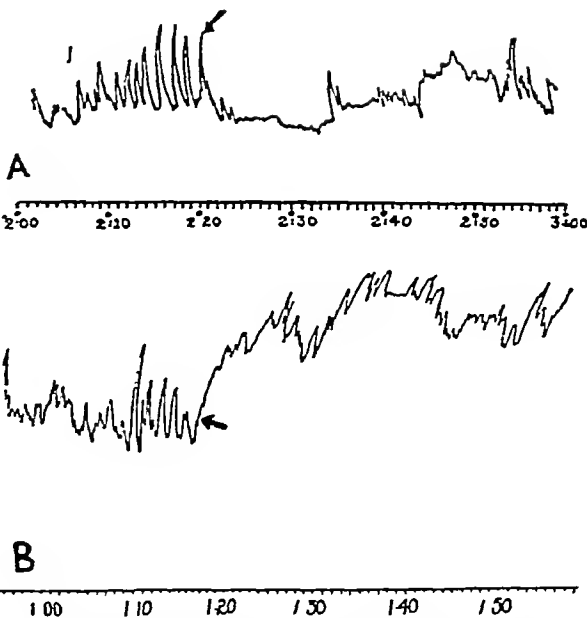


FIG 42.4 Showing normal (A) and abnormal (B) gastric feeding reflexes. The record B was from a patient with chronic appendicitis. Arrows indicate when food was given. Time (hour) given along the abscissae. (After Welch)

Some of the products of protein digestion in the intestine (e.g., the monoamino-monocarboxylic acids) and hydrochloric acid inhibit gastric tone and peristalsis. This action, which is referred to as the *enterogastric reflex*, is brought about through

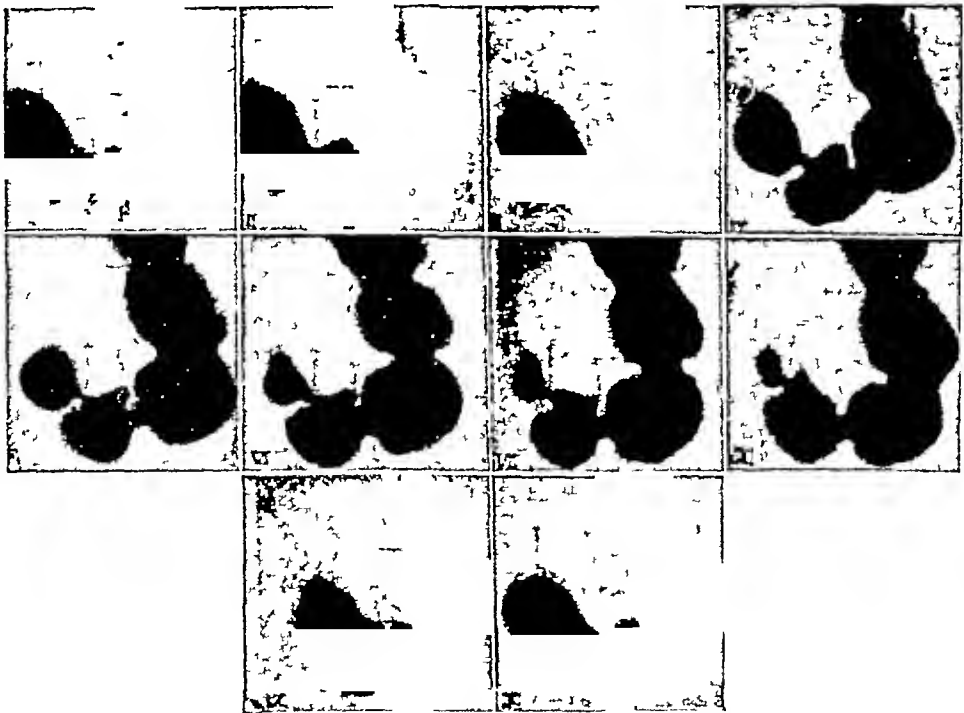


FIG 42.5 Serial X-ray photographs of human stomach taken at 2 second intervals during digestion. (From Cole)

the vagus Sugars in high concentration also inhibit gastric movements, this effect is probably induced through the medium of enterogastrone rather than by a reflex mechanism

In the intact normal stomach distension of the gastric walls acts as a stimulus though extreme distension is inhibitory Nervousness and anxiety tend to increase peristalsis while certain other emotional states—apprehension, fear, mental strain, shock or depression—tend, as a rule, to inhibit the movements and to reduce gastric tone Pain and physical fatigue act similarly Exercise unless very moderate tends to inhibit gastric motility After the operation of *va. omy* for peptic ulcer the motility of the stomach and the hyperemia of the mucosa caused by food are less than normal

It is believed that in many instances the discomfort or distress which follows shortly after a meal has its origin in one or other of such nervous factors as those just mentioned When food is taken while one is fatigued, anxious, agitated or hurried, disturbances of the normal motor mechanisms of the stomach give rise to unpleasant gastric sensations The precise manner in which these are set up is not clear, but it is not unlikely that, in some instances at least, with the inhibition of the normal descending waves of peristalsis, reverse waves arise which lead to heartburn, belching and a feeling of discomfort (p 561) But whatever the mechanism, the relation of psychic factors to gastric symptoms is clearly evident Thus so-called nervous dyspepsia is, to quote Alvarez, "the disease of the mother who prepares the meal and then wrangles with children or husband at the table, it is the disease of business men and women who gulp down some food at a counter and rush back to work, and it is the disease of the president of a luncheon club, or of the traveling salesman who gives 'pep talks' at luncheons and dinners"

Among chemicals which influence gastric motility the following may be mentioned Alcohol, lactic acid, coffee, sodium bicarbonate in therapeutic doses, histamine, insulin and cholinergic drugs, e.g., methylcholine chloride and carboaminocholine increase peristalsis, smoking, atropine, vitamin B₁ deficiency (p 750) infectious fevers and anoxia depress it Insulin is without effect upon motility after double vagotomy Ganglionic blocking drugs, such as tetraethylammonium chloride and the hexamethonium compounds, are powerful inhibitors of the gastric

movements Morphine increases the tone and amplitude of the contractions of the pyloric sphincter, its effect upon the antrum is variable

THE EMPTYING OF THE STOMACH

After an ordinary mixed meal the normal human stomach empties in from 3/4 to 4 1/2 hours A meal consisting mainly of carbohydrate leaves more rapidly than one containing much meat, and this more rapidly than one containing fat Sugar (glucose or sucrose) fed in concentrated solution retards evacuation In normal young adult males a test meal is evacuated in about 2 hours the range is from 1 5 and 3 3 hours (Van Liere and Northrup) Fluids and semi fluids commence to leave the stomach almost immediately after being swallowed (see table 40)

The theory propounded by Cannon, that the pylorus is controlled by the acidities of the gastric and duodenal contents—high gastric acidity causing opening and high duodenal acidity closure—is held no longer A stimulus of practically any sort arising in the duodenum can reflexly increase the tone of the pyloric sphincter and effect its closure, a high acidity can serve as such a stimulus, but it has no specific effect in this regard A high degree of alkalinity in the duodenum also inhibits gastric motility and delays evacuation The emptying time of the stomach is actually shortened by reducing the gastric acidity, as by the ingestion of sodium bicarbonate or of disodium phosphate (Van Liere and Sleeth) Bile salts reduce the emptying time of the human stomach Gastric evacuation is prolonged by general anesthetics, the greatest increase being caused by chloroform and ether and the least by nitrous oxide and cyclopropane

The rapidity with which raw egg white or water leaves the stomach, as well as the normal and sometimes shortened emptying time in gastric anacidity (p 518) had always offered obstacles which the acid theory of pyloric control found difficult to surmount McClure, Reynolds and Schwartz have shown clearly that no important relationship exists between the emptying time and gastric or duodenal acidity The maintenance of a constant acidity in the duodenum did not prevent pyloric opening On the other hand, when the duodenum was kept neutral or alkaline by means of fluid introduced through a duodenal tube, rhythmical opening and closure of the sphincter occurred in the usual way Nor were Baird

TABLE 40
Gastric evacuation—various foods
(After Wilson)

	MINUTES TO SWALLOW	FIRST LEAVING OF STOMACH	PERCENTAGE OUT OF STOMACH IN		
			1½ hours	3 hours	4½ hours
		minutes			
Carbohydrates					
Thick porridge	2	4	75	95	
Bread 40 grams, dates 100 grams	3	3	60	95	
Proteins					
Egg white, raw	1	3	75	85	90
Codfish, boiled	5	20	30	85	95
Lean meat, baked	5	7	40	80	90
Fats					
Cream 32 per cent	1	1	25	40	70
Bacon and egg yolk	5	12	10	30	95
Olive oil	1	1	25	50	60

and associates able to obtain evidence for the acid theory of pyloric control from experiments on human subjects. The emptying time of the stomach showed no dependence upon the reaction of the gastric or duodenal contents. Essentially similar results have been obtained by McCann and several other investigators.

The opinion is now widely held that the pylorus is patent for the greater part of the time and that the evacuation of the stomach is definitely related to the peristaltic activity of the antrum. Wheelon and Thomas obtained simultaneous records in animals by means of balloons placed, respectively, in the pyloric antrum and the pyloric canal, which showed that as the constricting ring traverses the antrum the sphincter becomes relaxed and the chyme is swept before the wave into the duodenum. The pylorus then closes for a moment, relaxing again and remaining open until the next wave of constriction arrives from the antrum. Graphic records obtained by means of balloons placed in the antrum combined with X-ray examination also show that the emissions of chyme through the sphincter are coincident with the passage of waves over the former region (Carlson and associates), and Baird and associates found that in the human subject fluids shortly after being drunk issued from a tube placed in the duodenum just beyond the pylorus not continuously, but in jets. The importance of the pylorus (the keeper of the gate) in controlling gastric evacuation has, therefore, been greatly exaggerated. X-ray examinations of the stomachs of patients in whom the pylorus has

been excised, show emptying times which do not differ significantly from the normal (Singleton).

It is apparent from the foregoing discussion that factors which increase or diminish *gastric tone* and the *force of the antral peristalsis*, will shorten or lengthen, respectively, the evacuation time. The immediate factor which determines the rate of gastric evacuation appears to be the pressure gradient between the antrum and the duodenal bulb. Fat, for example, inhibits gastric motility and so delays the emptying time.

The degree to which the gastric contents have been reduced to *fluid or semi-fluid consistency* appears also to be an important factor determining the rate of emptying of the stomach. As already mentioned, water drunk by the human subject passes through the pylorus in squirts almost immediately after having been drunk. In the early stages of digestion, when the food contains pieces of solid material, the tone of the pylorus is higher than in the later stages, when the gastric contents are in a semi-fluid state. Raw egg white and fluid milk also leave more rapidly than coagulated egg-albumin or clotted milk. The shorter emptying time of carbohydrate food as compared with protein is due most probably to the greater readiness with which the former is reduced to a semifluid state. The details of the mechanism whereby the consistency of the food and the emptying rate are correlated are not clear, unless, as has been suggested, solid particles act as mechanical stimuli which, coming into contact with the pylorus, cause pyloric closure and possibly set up retrograde

waves in the vestibule Cannon, for example, fed pellets of a bismuth salt to an animal and, observing the gastric movements by means of the X-ray, saw the opaque particles carried up to the pylorus by peristaltic waves They were not, however, allowed to pass into the duodenum but were returned to the vestibule They were retained in the stomach after the rest of the gastric contents had been discharged Retrograde waves in the vestibule have also been observed by Kleine and others when solid materials such as undigested meat came into contact with the mucosa in the region of the pylorus

The observations of Apperly and several other investigators indicate that the *osmotic pressure* of the gastric contents is another factor influencing the emptying time of the stomach Madeod and associates found that hypertonic glucose solutions when fed to rats were reduced to isotonicity before evacuation occurred, and McSwiney and Spurrell observed that hypertonic meals delayed gastric evacuation, proportionately to their hypertonicity Corresponding observations have been made by Van Liere.

Finally, *the state of the upper part of the duodenum* (p 572)—its fullness or emptiness, which thus alters the pressure gradient between the antrum and the duodenal bulb—has also a pronounced effect upon the state of the pyloric sphincter and the emptying time of the stomach It has been mentioned on page 566 that certain digestive products in the duodenum cause gastric inhibition through the enterogastric reflex or through the liberation of enterogastrone

The main factors, therefore, which influence the emptying time of the stomach are (a) *the motility of the stomach itself*, (b) *the consistency of the gastric contents*, (c) *the osmotic pressure of the gastric contents*, and (d) *the quantity and nature of the material in the duodenum* It should be remembered that the pressure in the antrum must exceed that in the duodenum in order for the stomach to empty Gastric evacuation is also influenced to some extent by the position of the body, emptying being more rapid when the subject is lying upon his right side than when standing or recumbent upon his left side

INTRAGASTRIC PRESSURES

The pressure within the human stomach varies in the standing or sitting position from 6 to 10 cm H₂O In

certain other positions, as shown by Wilson and Irving, e.g., lying on the right side, the pressure is frequently subatmospheric (−1 to −5 cm H₂O) The pressure in the fundus is considerably lower than that in the pyloric part. This indicates that the tonic contraction of the fundic walls does not serve to press the food towards the pylorus, as has sometimes been supposed The pressure is also usually higher in the antrum than in the duodenal bulb In the dog, intragastric pressures up to 90 cm of water have been observed The intra gastric pressure varies with the intra abdominal pressure and with the respiratory movements

As already mentioned, the entrance of material into the stomach causes a reduction in the tone of the gastric musculature, and an increase in the capacity of the stomach for the accommodation of the ingested food or drink (see postural tone, p 557) When, for example, a large quantity of water is drunk the upper surface of the fluid in the stomach as shown by X-ray rises very slowly and, after reaching a point about the middle of the body, remains practically stationary at this level, though an additional quantity of fluid be drunk, little change in the level of the fluid occurs Ordinarily, the intragastric pressure shows no increase after the ingestion of a meal, and even after a liter or so of water has been swallowed no rise in pressure, or only a very transitory one, can be detected Conversely when fluid is withdrawn from the stomach the gastric capacity becomes automatically reduced and the intragastric pressure remains virtually unaltered The adjustment of the gastric capacity to the volume of the food appears to be dependent in part at any rate upon a central reflex According to Kelling it is abolished by deep anesthesia Adaptation is shown, however, to some extent even by the excised surviving stomach The increase in gastric capacity is accompanied by reflex relaxation of the abdominal muscles If, on the other hand, an amount of fluid, which when taken into the dog's stomach causes no change in intragastric pressure, is injected into the peritoneal cavity, relaxation of the abdominal muscles does not occur and the intra abdominal pressure rises in proportion to the volume of the fluid which has been introduced The intragastric pressure rises, of course, to a corresponding degree The discomfort and sensation of fullness after a meal eaten hastily or under conditions of nervous strain are attributable in many instances to the inability of the gastric wall and abdominal muscles to become accommodated rapidly enough to the increased volume of the stomach contents

THE INNERVATION OF THE GASTRIC MOVEMENTS

The stomach contains intrinsic plexuses (of Auerbach and Meissner) similar to those in the intestine (p 583) Its extrinsic innervation is through both the right and left vagus nerves and the sympathetic (fig 42 7) The fibers of the vagus

end around cells in Auerbach's plexus.² The post-ganglionic fibers of the sympathetic (cell stations in celiac ganglion) end by arborizations around the muscle cells. The effects of the experimental stimulation of either of these nerves cannot always be predicted. Vagal stimulation may cause either augmentation or inhibition of gastric motility. The result obtained depends upon several conditions, e.g., the frequency and strength of stimulus, the level of the tone of the muscle at the time and the activity of the peristaltic movements. Generally speaking, when the tone of the gastric muscle is high and actively contracting, inhibitory effects are produced by vagal excitation, whereas when the muscle shows little or no motility and its tone is low a motor response is obtained from the vagus. Thus, McCrae and McSwiney, and Stopford found that when the *rhythmical movements* were present in the antrum, vagal stimulation caused inhibition but, if the movements were absent, vagal stimulation initiated them. Similarly in the case of the sympathetic, when hypotonus or hypertonus pre-exists, stimulation causes augmentation or inhibition, respectively, of gastric motility. Inhibition of the rhythmical movements of the antrum, however, is the frequent result of sympathetic stimulation. The sympathetic is generally considered to be excitatory to the pyloric sphincter and the vagus inhibitory, but an opposite effect may be obtained from either nerve, the tonus level at the time appearing to be a determining factor. Though variability of response is quite evidently a characteristic of the gastric nerves, their functions can be summarized in the statement that *the vagus is predominantly motor, the sympathetic predominantly inhibitory* with opposite effects, respectively, upon the pyloric sphincter. Thomas and Crider observed that, in dogs, vigorous gastric peristalsis ceased immediately upon sectioning the vagus nerves, the stomach remained quiescent for hours. But when the sympathetic nerves are sectioned as well, although there is inhibition of peristalsis for a time, motility soon

² Both vagus nerves enter into the formation of the esophageal plexus. From the latter, two trunks, each of which contains fibers from the right and the left vagus, emerge. The anterior vagal trunk supplies the anterior surface of the stomach and the posterior trunk the posterior surface. Each aspect of the stomach thus receives fibers from both vagus nerves. The pylorus and first part of the duodenum receive a special branch which arises from the anterior trunk. This branch also sends twigs to the liver. The posterior trunk supplies a large branch to the celiac plexus. (See McCrae.)

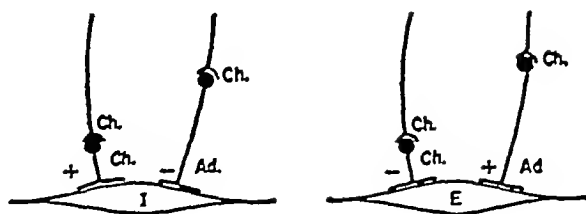


FIG 42.6 Diagram of innervation of the two main types of smooth muscle classified on the basis of effects of adrenaline and acetylcholine. In each of the two diagrams the parasympathetic innervation is shown on the left and the sympathetic innervation on the right. The type of smooth muscle on the left is inhibited by adrenaline and contracted by acetylcholine, the type on the right is contracted by adrenaline and inhibited by acetylcholine. E and I designate two types of "receptive mechanisms" in the cells which must be postulated to explain the fact that a single chemical, e.g. adrenaline, causes one cell to relax and the other to contract. (From Youmans, Amer Jour Med Vol 13, p 213, 1952.)

returns. This suggests that a tonic inhibitory influence, exerted by the sympathetic innervation is responsible for the abolition of gastric motility when the vagi alone are sectioned. In man, decrease in tone and in the motility of the stomach and reduction in the tone of the pyloric sphincter have been reported to result from section of the gastric vagi. On the other hand, section of the sympathetic innervation has been found to increase the motility of the human stomach. The innervation of the cardia is discussed on page 561.

After section of both sets of extrinsic nerves to the stomach, the movements when they return are dependent upon the intrinsic nervous mechanisms. The evidence for the mediation of excitatory effects of the vagus by acetylcholine is given in ch. 72. The gastric vagi also contain afferent fibers. Section of one vagus and stimulation of its central end is followed by movements of the stomach.

In order to explain excitation of the gastric wall together with relaxation of the sphincter by vagal stimulation and converse effects by sympathetic stimulation, the existence of two functionally distinct receptive mechanisms in each muscular site has been proposed, rather than that the opposing effects of the vagus or sympathetic are mediated by two different types of fiber contained in each of the two nerves (see Youmans). One mechanism (in muscle of gastric wall) is excited by acetylcholine and inhibited by adrenaline

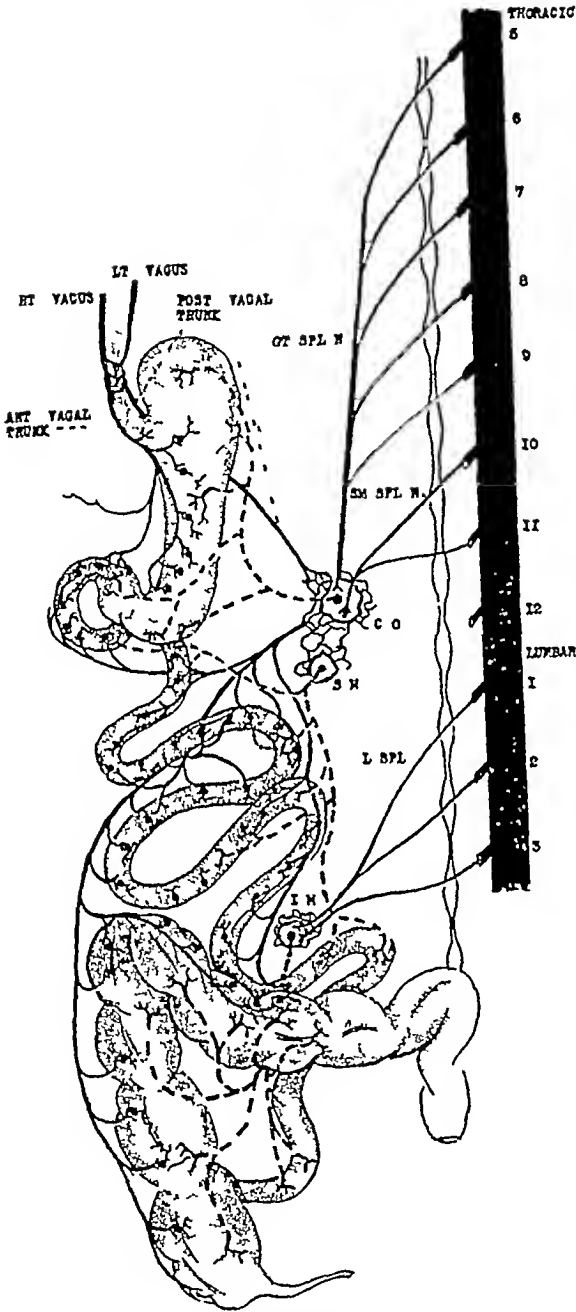


FIG 42.7 Diagram of the innervation of the stomach, small intestine and proximal part of the colon. GT SPLN, great splanchnic nerve, SM SPLN, small splanchnic nerve, L SPL, least splanchnic nerve, CG, celiac ganglion, SM, superior mesenteric ganglion, IM, inferior mesenteric ganglion. Continuous lines, vagal and sympathetic preganglionic fibers, broken lines, sympathetic postganglionic fibers, ganglion cells and postganglionic fibers of vagus in gastro-intestinal wall.

(fig 42.6) The other (in sphincter muscle) is inhibited by acetylcholine and excited by adrenaline. Thus stimulation of the cholinergic nerves of the vagus causes excitation of the mural muscle and relaxation of the sphincter, stimulation of adrenergic fibers of the sympathetic causes relaxation of the mural muscle and contraction of the sphincter.

Evidence for a diencephalic center for gastric movements has been obtained by Beattie and Sheehan. Stimulation of the hypothalamus in the region of the tuber cinereum resulted in contractions of the stomach, the effect was abolished by section of the vagi. Stimulation of the posterior hypothalamic region, on the other hand, inhibited all gastric movements. Sheehan in a later communication reports that cooling or faradic stimulation of the frontal or premotor area of the cortex in the monkey inhibited the movements of the stomach which had been contracting actively. Contractions could not, however, be initiated in a resting stomach by stimulation of the cortex.

ACUTE DILATATION OF THE STOMACH

In this grave though comparatively rare condition the stomach becomes greatly, sometimes enormously, distended, its cavity contains large quantities of fluid and gas which may stretch its walls to paper thinness. The accumulated fluid of course represents a corresponding loss of water and chloride from the body's stores. Reduction in blood chloride, alkalosis and dehydration result (p 23), tetany may occur (p 852). There is profound prostration, collapse ending in death is not uncommon. The condition, in the majority of instances, occurs as a sequence to an abdominal or pelvic operation but it occasionally follows a blow upon the abdomen, some severe injury or acute illness, childbirth, a drinking bout or overeating.

The condition is apparently of reflex origin, the initial and essential factor being the loss of gastric tone and peristalsis. The stomach then becomes distended by the accumulation of its own secretions which it is unable to expel, as well as by the regurgitation of fluids from the duodenum. There is no obstruction at the pylorus. Even the normal stomach and also probably the upper part of the duodenum cannot absorb the fluids which they secrete. Therefore, as a result of the paralysis of the gastric wall gas and secretions simply collect and distend the viscus. The condition would thus appear to be analogous to paralytic ileus (p 594). Hypersecretion, apparently, is not a factor in the early stages at any rate. Later, however, the distension of the gastric walls probably acts as a stimulus to secretion (p 508). In some instances, swallowed air and intestinal gases play a part in ballooning the relaxed organ. Ac-

cording to Dragstedt when the dilatation reaches a certain degree an added factor comes into play, the enlarged stomach by its pressure obstructs the third part of the duodenum.

THE DUODENAL BULB OR CAP

The first two inches of the duodenum are referred to as the duodenal bulb or cap, this region is of special interest since it is the site of duodenal ulcer (p 520). Functionally, the duodenal bulb is considered by most observers as part of the stomach. It is somewhat triangular in outline with its base surmounting the pylorus and is directed backwards, upwards and to the right (fig 42.2). Its glands (Brunner's) secrete an alkaline fluid. Its junction with the duodenum proper is marked by a slight thickening of the circular muscle fibers which are believed by some to exert a sphincter-like action. During the period of gastric evacuation the bulb becomes filled with chyme. From time to time a peristaltic wave sweeps over it from base to apex, carrying its contents into the next section of the duodenum. The chyme is also conveyed into the duodenum proper in a passive manner, that is, by the simple overflowing of the bulb. Occasionally an antiperistaltic wave arises in the duodenum below the cap, continues upwards to the pylorus and appears in the pyloric vestibule.

Related movements of the pyloric vestibule, pyloric sphincter and the duodenum below the bulb

Wheelon and Thomas placed recording instruments in the pyloric antrum, pyloric canal, and duodenum distal to the cap, and obtained simultaneous tracings of the contractions in these three situations. The movements of the antrum and sphincter were found to be related with those of the duodenum. During the contraction of the antrum the duodenum is relaxed, but about one minute after the commencement of the relaxation of the former the duodenum commences to contract. That is, relaxations of the antrum roughly coincide with contractions of the duodenum and contractions of the antrum with relaxations of the duodenum (receptive relaxation of the duodenum). Contractions of the sphincter were found to occur rhythmically from 3 to 5 times per minute, each contraction commencing while the duodenum was relaxed. The contraction of the duodenum commenced $2\frac{1}{2}$ seconds later but the contractions in the two situations reached their maximal heights at about the same time. These movements of stomach and duodenum are correlated apparently through the intrinsic plexuses in the gastrointestinal wall. A connective tissue barrier exists between the muscle of the stomach and duodenum.

which must prevent the continuous spread of the wave independently of a nervous mechanism of some sort.

Movements initiated in the duodenum have been shown to influence the activities of the pyloric sphincter. When the duodenal mucosa was stimulated an ordinary peristaltic wave appeared which travelled down the bowel in the usual way (p 578), but a firm contraction of the sphincter also occurred, followed by a prolonged period of relaxation. A series of such stimuli produced a continuous contraction of the pyloric sphincter. Activity of the duodenum has been shown by several observers to have the effect of increasing the tone of the sphincter. The state of the duodenum in this way affects indirectly the emptying of the stomach, and the filling and emptying of the cap. When the duodenum is filled with chyme active movements are set up in its wall which have the two-fold effect of withdrawing material from the cap and causing reflex closure of the sphincter. When the duodenum becomes empty its activity subsides, the pylorus relaxes more fully and the emptying of the stomach is hastened. This effect of a full duodenum upon the emptying time of the stomach is shown by the fact that a meal taken after a period of fasting is discharged much more rapidly than if the duodenum is already replete with chyme, gastric evacuation is then delayed. Distension of the bowel by a balloon or its irritation by chemical or mechanical means increases pyloric tone. It has been shown repeatedly on the other hand that if the chyme as it issues from the pylorus is allowed to escape through a fistulous opening instead of being permitted to fill the duodenum, the stomach empties with greater speed. It has been found by Thomas, Crider and Mogan that a reflex effect of even greater importance than that upon the sphincter is exerted from the duodenum upon gastric peristalsis. Substances introduced into the duodenum markedly reduced the force of the peristaltic waves in the antrum, whereas draining the duodenum increased gastric motility and hastened gastric evacuation.

Antiperistalsis in the second and third parts of the duodenum

Antiperistalsis is a normal occurrence in the duodenum below the cap, the reverse waves can be seen during X ray examination of the human subject passing orally and conveying material to a higher level of the duodenum, into the cap or through the pylorus into the stomach.

VOMITING OR EMESIS

The mechanism of vomiting involves the co-ordinated actions of the muscles of the stomach, esophagus, and abdominal wall. The act may also be associated with antiperistaltic movements in the intestine. The muscular mechanisms are gov-

erned by a center in the medulla, which discharges impulses along numerous efferent nerves, and are influenced by afferent impulses arising in the stomach, in other viscera or in practically any region of the body. Or, the center may be excited by substances conveyed in the blood stream (p 575).

THE VOMITING MOVEMENTS

Cannon has given a graphic account of the act of vomiting in the cat. The movements were observed by means of radiography after a bismuth meal and the administration of apomorphine, which excites the vomiting center. The upper part of the stomach showed complete inhibition of its tone and appeared as a perfectly flaccid bag, the cardia relaxed. There then followed several deep peristaltic contractions which, commencing about the middle of the body of the organ, swept downwards toward the incisura angularis where they came to a standstill forming a sharp ring of constriction. From this point a weaker wave continued to the pylorus. Finally, a very deep, strong contraction at the incisura appeared to almost divide the stomach in two, the upper part of the stomach and the cardia meanwhile remaining quite relaxed. A sharp contraction of the diaphragm and abdominal muscles then followed and ejected the gastric contents through the open cardia into the esophagus. The stomach played a more or less passive part in the process,¹ its evacuation being effected by the strong compression to which it was subjected by the sharp descent of the diaphragm and the contraction of the abdominal muscles. Antiperistalsis was observed only once and then the wave did not proceed beyond the antrum. The deep contraction at the incisura offered an effectual barrier to the passage of stomach contents in a downward direction.

During the ejection of the vomitus the esophagus is relaxed throughout, the glottis is closed and the respirations are inhibited, the larynx and hyoid bone are drawn upward and forward and are held rigidly in this position. The throat is thus enlarged to allow free exit for the stomach contents which are prevented from entering the naso-pharynx by the evaluation of the soft palate.

Similar movements have been described in man.

¹ Magendie showed long ago that an active gastric element was not necessary for the vomiting act, since its essential features could be induced in animals after the stomach had been replaced by a pig's bladder.

Definite antiperistaltic waves in the stomach are rarely seen, though violent churning movements may occur⁴ According to Barclay, just prior to the commencement of the vomiting movements, a sudden reduction of gastric tone occurs, the lower limit of the stomach, as seen radiographically, dropping a couple of inches This coincides with the sensation of nausea (p 602) which ordinarily precedes vomiting Loss of tone and reduced motility of the jejunum has also been observed in animals preceding the emesis caused by apomorphine In certain types of vomiting, e g , intestinal obstruction, or in persistent vomiting from other causes, antiperistalsis arises in the small intestine and sweeps material into the stomach or there may be a strong contraction of the duodenum which reverses the pressure gradient between the antrum and the duodenal bulb Such movements of the duodenum occur sometime prior to the actual vomiting or at the same time, it accounts for the fact that a short time after the stomach has been thoroughly washed out, bile-stained fluid or fecal material may be vomited According to Alvarez, reverse peristalsis starting in the upper bowel is itself a potent cause of nausea and vomiting

Relaxation of the cardia is an essential part of the vomiting act, for the stomach is subject to strong compression during coughing, defecation, etc , yet the gastric contents are not as a rule forced into the esophagus The tone of the cardia is probably actually increased at these times It has been mentioned that division of the vagi in animals causes the cardia to enter into a spastic state Hatcher and Weiss found that if after such a procedure a vomiting reflex was initiated, mucus was expelled from the esophagus and the usual muscular movements were called into play, with the exception of relaxation of the cardia That is, the animal retched but material was not expelled from the stomach It is well known that with some persons vomiting is difficult while in others little distress is experienced Differences in the degree of

⁴ In certain lower forms, however, e g , the fish and the frog, which of course have no diaphragm, vomiting is carried out by the activity of the stomach alone, antiperistaltic waves carrying the food through the cardia This more primitive type of vomiting, i e , where reverse peristalsis and relaxation of the cardia are the prominent features, occurs normally in infants The excess fluid of an oversized meal is regurgitated without the assistance, apparently, of the abdominal muscles or diaphragm, position and external pressure upon the abdomen following the meal sometimes, no doubt, also play a part.

TABLE 41

DRUGS	AFFERENT AUTONOMIC FIBERS ACTED UPON	ORGANS CONTAINING SUSCEPTIBLE FIBERS
Mercuric chloride	Sympathetic and vagus	Stomach
Tartar emetic	Vagus	Stomach, duodenum and heart
Digitalis	Sympathetic and vagus	Heart
Pilocarpine	Sympathetic and vagus	Heart

tone of the cardia are probably responsible for these individual peculiarities (table 41)

THE VOMITING CENTER

The vomiting center lies in the dorsal part of the lateral reticular formation of the medulla oblongata close to, and to some extent including the tractus solitarius It lies in close relation to, but is quite distinct from the respiratory centers It is one of a constellation of visceral centers in this situation—salivatory nuclei, defecation and vasomotor centers, as well as the vestibular nuclei—a relationship consistent with the physiological reactions (salivation, vasomotor and respiratory, etc) associated with the vomiting act (see also p 575) The *efferent* fibers are contained chiefly in the phrenics, the vagi and the sympathetics, but fibers are also conveyed by spinal nerves to the abdominal muscles and by cranial nerves to the muscles of the pharynx, palate, etc The *afferent* impulses reach the center along a multitude of routes, the chief being the vagal and sympathetic fibers of the stomach and abdominal viscera

THE INDUCTION OF VOMITING

The vomiting center may be excited (a) reflexly by impulses arising in the stomach, or some other part of the body, (b) by impulses received from cerebral centers, or (c) by chemical materials carried to it in the blood stream

Vomiting of reflex and psychic origins

Mechanical and chemical irritants of various kinds act upon vagal or sympathetic afferent terminals in the gastric mucosa and initiate the vomiting reflex Toxic materials taken in the food or formed during digestion may act in this way Among the specific excitants of the vomiting act—

emetics—are tartarate of antimony (tartar emetic), mustard, ipecac, mercuric chloride, copper and zinc sulphates, etc. Some of these have a selective action upon the terminations of one or other of the two groups of afferent autonomic fibers. For instance, tartar emetic, in the stomach, acts through the vagus but not through the sympathetic, since it fails to act after vagotomy, or after the vagal effect has been annulled by atropine. Mercuric chloride, on the other hand, acts more especially on the sympathetic mechanism since atropine does not prevent its emetic action, but ergotoxine which acts specifically by paralyzing sympathetic fibers, does. Mustard, salt and water, copper sulphate and zinc sulphate act upon both types of nerve.

Impulses arising from structures other than the stomach, particularly from other parts of the alimentary canal, are also powerful excitants of the vomiting center. It is well known that even mild mechanical stimulation of the pharynx or fauces is effective. Inflammation or mechanical disturbances in the intestine, e.g., appendicitis, obstruction, strangulation, etc., may induce violent emesis. Distension of the duodenum or jejunum in a conscious animal induces vomiting movements. Impulses arising in the kidney, bladder, uterus, gall-bladder, heart or any other viscus may induce vomiting quite independently of any condition within the stomach itself. Hatcher and Weiss have shown that several substances cause emesis through their action upon the heart, e.g., digitalis in large doses and pilocarpine. These drugs act apparently upon both vagal and sympathetic mechanisms. Tartar emetic also acts through the cardiac vagus as well as through the gastric nerves. These facts afford an explanation of the nausea and vomiting that occur in heart failure and in disease of the coronary vessels.

Severe pain, wherever located, strong emotion or impulses arising in the organs of special sense, e.g., those of smell, taste or sight, may cause vomiting. In the dog a *conditioned vomiting reflex* may be readily established by morphine injections (ch. 69).

Motion sickness

The equilibratory sense organs of the labyrinth—otolith organs and semicircular canals—are stimulated by linear acceleration or by angular acceleration as in rotational swaying or rolling motions. Labyrinthectomy or section of the vestibular nerves abolishes an animal's susceptibility to

motion sickness, and deaf-mutes (in the dark or blindfolded) whose labyrinths are undeveloped, or infants who have not commenced to walk and whose equilibrating mechanisms are as yet not functioning fully, do not suffer from car or seasickness. Air sickness is caused most commonly by linear acceleration in a vertical direction, as when sudden changes in air density cause the plane to rise or fall. Ascent or descent in an elevator (especially descent), and swing motions are other examples of vertical linear acceleration which cause a sickening sensation (see also chapter 65).

Seasickness is also due chiefly to stimulation of the otolith organs by vertical linear acceleration, as when the ship pitches, rolling motions appear to exert a less pronounced effect. In any case the impulses are set up in the otolith organs rather than in the semicircular canals, for the angular acceleration does not reach the threshold required for stimulation of the cristae, nystagmus does not occur. Of the two sets of otolith organs—utricle and saccule—the utricle, not the saccule, seems to be the one responsible. The reflexes set up through the vomiting center and the autonomic centers are mainly responsible for the symptoms of seasickness and other forms of motion sickness. No significant changes in blood chemistry have been found, unless secondarily following a prolonged period of vomiting.

The susceptibility to seasickness is usually much reduced in the recumbent position, and some persons who show great distress when erect become quite comfortable when they lie down, especially when the head is inclined backward. In this position the receptors are removed from the plane of the motion. It is also well known that after a few days at sea most persons become accustomed to the motion of the ship, and do not suffer from seasickness. The mechanism whereby this tolerance is developed is obscure.

It has been shown by Bard and his associates that excision of the nodule of the cerebellum (ch. 70) completely prevents motion sickness (swinging) in dogs, but not vomiting from other causes. Ablation of no other part of the cerebellum or of any portion of the cerebral cortex had this protective action.

Though vestibular stimulation is by far the most potent factor causing motion sickness, impressions through the eyes are often contributory. For example, viewing the motion picture of a ride in a scenic railway, which would cause sickness if

actually experienced, may give rise to nausea. Smells, or various sensations from the gastrointestinal tract are well known to increase the susceptibility to seasickness.

During the second World War intensive research was undertaken to ascertain the best means of preventing or ameliorating motion sickness, a large number of drugs were investigated. Among those found to give the greatest degree of protection are, *mosidal* (ethyl- β -methyl-allyl-thio-barbituric acid), which was found to be highly effective in preventing motion sickness (swinging) in dogs, *hyoscine hydrobromide*, *hyoscyamine hydrobromide*, and *benzedrine*. A combination of effective drugs usually gives greater relief than one alone.⁵ The most recent and most promising addition to the list of motion-sickness remedies is a benadryl derivative called *dramamine* (β -dimethylamino-ethyl benzohydroxyether-8 chlorothiophyllinate). This drug was found highly effective both as a preventive and as a cure in tests carried out on naval personnel by Gay and Carlner, and on airmen by Strickland and Hahn.

"CENTRAL" VOMITING

Certain substances such as *apomorphine*, *ipecac* (or *emetine*), *picrotoxine*, when administered intravenously cause vomiting through a specific action upon the center. These substances induce vomiting in animals when applied directly to the center. Typical vomiting efforts (retching) follow the injection of apomorphine after removal of the stomach and intestines. The experiments of Wang and Borison point to a chemoreceptor "trigger" mechanism in the *ala cinerea* which, through its connection with the vomiting center itself, induces vomiting when stimulated by these emetics. Copper sulfate was found by these investigators to act as an emetic both peripherally, through the gastro-intestinal tract, and centrally a lesion in the *ala cinerea* abolished the central action only.

General bodily conditions, e.g., mental stress, fatigue or toxic substances carried in the blood stream, may induce emesis either directly or possibly by raising the excitability (lowering the threshold of stimulation) of the nerve cells of the center so that afferent impulses from various

sources which ordinarily are of subthreshold value become effective.⁶

It is evident then that vomiting does not necessarily imply a primary disturbance of the gastrointestinal tract.

The persistence of pernicious vomiting of pregnancy is probably due to increased excitability of the center resulting from some metabolic disturbance. According to Harding the important factors in its production are carbohydrate starvation and dehydration with ketosis. Ordinary morning sickness is looked upon as a minor form of this condition, and differing from pernicious vomiting only in degree. As a result of the demands of the fetus, liver glycogen is low during pregnancy. After the fast of the night the reserves of carbohydrate are further reduced, a mild ketosis results which leads to nausea and vomiting. The distaste for food by hindering the replenishment of the carbohydrate stores aggravates the ketosis. Thus morning sickness merges insensibly into the pernicious type of vomiting. A neurotic element is also frequently a potent influence in the development and persistence of the condition. The obvious corrective measures are, high carbohydrate feeding, the free administration of fluids, injections of glucose if necessary, mild sedatives, rest and quiet.

Interference with the blood supply to the center, as by elevation of the intracranial pressure through tumor, etc., or by anemia or hemorrhage, may lower its threshold to the point where vomiting ensues. Vomiting associated with raised intracranial pressure occurs suddenly, is forceful and not preceded by nausea or retching (*projectile vomiting*). Vomiting may be induced in animals by tying the carotid and vertebral arteries. The vomiting of mountain sickness is also a phenomenon associated with anoxemia (p. 422).

THE MOVEMENTS OF THE STOMACH IN HUNGER

The movements of the stomach during fasting have been studied by a number of observers, including Cannon and Washburn, and Carlson. Carlson, in an exhaustive series of experiments, has investigated the contractions and tonus changes of the empty stomach in various species of animals and in man. The movements were recorded by means of a thin rubber balloon introduced into the stomach and connected by tubing.

⁶ It is of interest in this regard that substances formed in the body, e.g., adrenaline, choline, and histamine induce vomiting when applied directly to the center.

⁵ The Canadian Motion Sickness Remedy—National Research Council Formula—is

Hyoscine HBr	0.1 mg
Hyoscyamine HBr	0.3 mg
Ethyl- β methyl-allyl barbituric acid (Mosidal, V ₁₂)	130 mg

with a manometer containing water, chloroform or bromoform. A float placed upon the surface of the fluid carried a writing point by means of which the contractions were recorded upon a moving surface (fig 42 8)

In man the movements of the empty stomach are

of two types (a) a *tonus rhythm* of the fundus and body and (b) the *lunger contractions* (fig 42 9)

The tonus rhythm

The tone of the fundus and body of the stomach increases steadily as digestion progresses. When

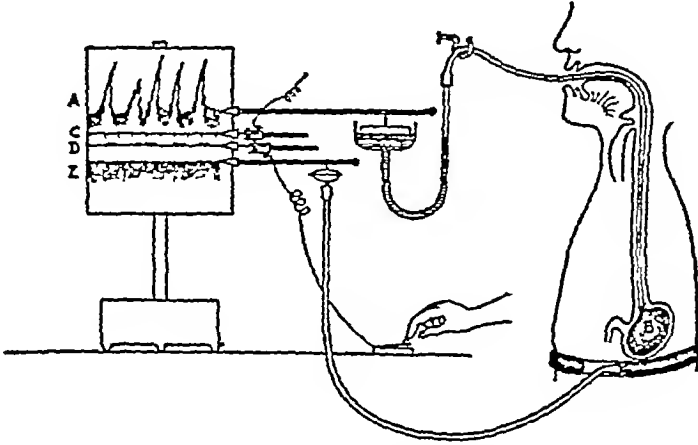


FIG 42 8 Diagram showing the method used to record the gastric hunger contractions. A, kymograph record of the increase and decrease of volume of the gastric balloon B, C, time records in minutes, D, record of the subjective experience of hunger pangs, E, record of the pneumograph placed about the waist. (From Cannon.)

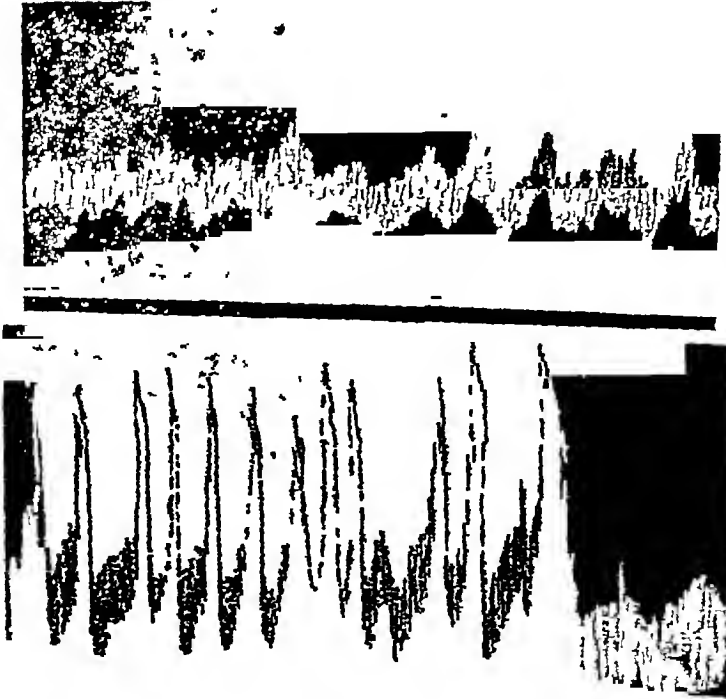


FIG 42 9 Upper, tracing of the tonus rhythm of the stomach (man) three hours after a meal. Lower, tracing from the stomach during the culmination of a period of vigorous gastric hunger contractions (From Carlson.)

the stomach has been nearly emptied the tone is near its maximum, and shows rhythmical fluctuations. The tonus changes are never very marked features and require a rather delicate recording instrument to make them evident. They were not recorded in the earlier experiments of Cannon and Washburn. They have been observed directly by Carlson through a gastric fistula in man when the stomach was practically empty of food and contained no balloon or other foreign material. The rate of the tonus waves is slow, but very regular. They have a frequency of about 3 or less per minute.

The hunger contractions

These are powerful peristaltic waves which arise in the cardia and sweep over the stomach. They commence about three hours after a meal, i.e., when the stomach is nearly empty, and are superimposed upon the tonus rhythm. The hunger contractions do not occur continuously, but in series (hunger periods) separated by intervals in which the stomach, except for the tonus rhythm, is quiescent. The duration of each hunger period is usually from 30 to 45 minutes, though it may continue for $1\frac{1}{2}$ hours or be as short as 6 minutes.

The longer periods are found when the tonus of the fundus is high, the shorter ones are associated with a low tonus. The quiescent intervals last from $\frac{1}{2}$ to $2\frac{1}{2}$ hours. The duration of individual contractions is from

20 to 25 seconds, and their amplitude varies throughout the series. They are relatively feeble at the commencement, but become of maximal strength about the middle of the hunger period and die down again toward the end. Their frequency increases progressively throughout. At the beginning of the period they occur several minutes apart, but toward the end they succeed one another rapidly so that the intervening pauses are practically abolished and the muscle may pass into incomplete tetanus which usually lasts for from 2 to 5 minutes, but may be as long as 15 minutes. It occurs especially in young robust individuals, being unusual in older persons. Discomfort or actual pain (hunger pang) is associated with the contractions. In infants the tonus rhythm and the hunger contractions have been demonstrated shortly after birth and before the stomach has been filled. When the contractions become well marked the infant wakes and cries.

During a fasting period in the human subject investigated by Carlson extending over a five day period the hunger contractions showed no diminution. They, as well as the tonus rhythm, actually increased in amplitude, yet the hunger pangs and the general sensation of hunger became less after the third day. It is well known that the subjects of prolonged fasts experience little discomfort after the first few days and even the desire for food may pass.

The hunger contractions were found by Carlson to be inhibited by various means, e.g., smoking, compression of the abdomen, as by "tightening the belt", taking a quantity of water into the stomach, strenuous muscular exercise, and the application of cold to the surface of the body.

MOVEMENTS OF THE ALIMENTARY CANAL (*Continued*), ABSORPTION FROM THE INTESTINES, THE FORMATION OF FECES, CONSTIPATION, INTESTINAL OBSTRUCTION

MOVEMENTS OF THE SMALL INTESTINE

Three types of muscular movements are usually described as occurring in the small intestine (a) *peristalsis*, (b) *segmenting contractions*, and (c) *pendular movements*

PERISTALSIS

This type of movement has already been mentioned as occurring in the esophagus and stomach. From a study of the peristaltic wave in the bowel Bayliss and Starling formulated the "*law of the intestine*", which states that a stimulus applied to a given point in the intestinal wall initiates a band of constriction on the proximal side and relaxation on the distal side of the stimulated point. The two phases, contraction and relaxation—excitation and inhibition—were described as travelling down the bowel at the same rate, constituting a wave which swept the intestinal contents before it. It is questionable, however, whether relaxation of the bowel on the distal side of the stimulated point actually occurs. Henderson was unable to demonstrate it in the intestine of either the guinea-pig or rabbit, a simple constricting ring being the only movement observed. The natural stimulus to peristalsis and other types of intestinal movement is the distension of the intestinal wall caused by the food mass within its lumen. Quigley and associates have shown that a bolus is propelled along the bowel in a spiral fashion. They inserted a bolus of cotton to which two long threads of different colors were attached and found that the rotation occurred in an anticlockwise direction. The length of bowel traversed in making a full rotation (360°) was from 23 to 35 cm. The bolus was propelled along the intestine (ileum) at an average rate of 9 cm. in 8 or 9 minutes.

It has been mentioned that waves may be seen passing in an oral direction (antiperistalsis) in the duodenum (p. 572). For a variable distance above the ileocolic valve antiperistalsis also occurs, and appears to serve as a check to the too rapid passage of ileal contents into the caecum. With these exceptions, peristalsis normally travels only in an aboral direction. That the small intestine for the

most part is incapable of transmitting peristaltic waves in the opposite direction was demonstrated by Mall. He resected a segment of the small bowel and then restored the continuity of the tube by suturing the resected portion in the reverse position. Peristaltic waves travelled from below upwards in the reversed loop, and obstruction occurred due to the accumulation of food at the upper suture line. Cannon in a similar experiment also observed that the reversed loop offered a barrier to the passage of the food. He saw, in addition, the food carried toward the pylorus in the normal segment of intestine above the loop. This latter observation is in accord with the belief that reversed peristalsis occurs in pathological conditions, e.g., intestinal obstruction, and is responsible for the passage of fecal material into the stomach. It should be mentioned, however, that if the reversed loop be quite short peristaltic waves arising in the normal bowel above may simply push material through it as though it were an inert tube. Also it has been reported that if the animal survives the operation for a sufficient length of time (several weeks), the reversed segment may adapt itself to the altered conditions and move the food downward by true peristalsis (i.e., antiperistalsis in so far as the inherent ability of the bowel is concerned).

In animals (dogs) examined by means of X-rays, peristalsis is not as a rule of frequent occurrence during fasting and when waves do appear they follow, not in a regular series, but at odd intervals. Within a minute or two after eating, peristaltic waves appear in the duodenum and jejunum which previously had been quiescent. The peristaltic movements are of two types. *First*, a slow, gentle wave which moves sluggishly along the bowel at a rate of from 1 to 2 cm. or less per minute, may be seen from time to time. This transfers food masses, which have already been subjected to the more active segmenting movements, for short stretches along the intestine. *Secondly*, a swiftly moving peristaltic wave appears from time to time and coursing down the bowel for longer or shorter distances, sweeps all before it and then dies out.

The food is left for a later wave to carry it further, the food is thus conveyed along the bowel in relays. This type of movement is spoken of as the "*peristaltic rush*" Its speed may be anywhere from 2 to 25 cm per second, the farther it goes the greater speed it attains, the average rate is about 10 cm per second. The length of bowel traversed by any wave varies in accordance with the general state of activity of the bowel at the time and with the strength of the stimulus. After a strong cathartic, or as the result of some intense stimulation of the gastrointestinal tract as by an irritant poison, for example, the wave may sweep with great rapidity from pylorus to anus, completing the entire journey in a minute or less.

Experiments made upon animals show that a rush may be initiated by conditions within the stomach or even by an act of swallowing, and one of the most effective means of producing such a movement in the intestine is to have the animal drink. In some instances the rush is evidently started by the passage of material through the pylorus into the duodenum, but in many others, though a peristaltic wave appears in the stomach this does not reach the pylorus before a rush commences in the duodenum. This fact, and the observation that a swallowing act alone will start a rush indicates that in many instances the mechanism is reflex in nature. A peristaltic rush in the small intestine is frequently followed by a similar though independent movement (mass peristalsis) in the colon (p. 581).

In the human duodenum, following a barium meal, peristalsis is not a prominent feature. The duodenal contents are, nevertheless, constantly moved onwards. Barclay suggests that this is due to contraction of the muscularis mucosae, the mucous membrane being raised into transverse folds which ripple along the bowel.

THE SEGMENTING MOVEMENTS

When the abdomen of an animal is opened and the coils of intestine floated upon a warm saline bath in order to obviate the effects of cooling and drying, rhythmical contractions of the bowel wall may be seen. The movements consist of simple constrictions of the tube, resulting from sharp contractions of the circular coat of the bowel. The annular constrictions, which are analogous with the rhythmical contractions described as occurring in the gastric antrum, do not progress along the bowel, and cause little or no propulsion of the food.

According to Cannon the segmenting movements in an animal given an opaque meal and examined radioscopically appear in groups of simultaneous constrictions spaced regularly along the bowel. The groups succeed one another rhythmically at the rate of about 20 per minute. They divide the intestinal contents into numerous small sections which are redivided by the next group which follows. The halves of adjacent segments so divided flow together to form fresh masses which are in turn bisected and reformed by the fusion of the divided parts. This process is repeated over and over again and the rhythmical kneading of the food results in a thorough mixture of its constituents with the digestive juices. It also encourages the absorption processes by continually bringing the chyme into contact with fresh mucous surfaces, and by the massaging effect of the movements upon the bowel wall, the flow of lymph and blood through the lacteals and capillaries is hastened. The segmenting movements have been observed by Hertz and others in the human intestine by radioscopes but their rhythm was slower—about 8 per minute.

PENDULAR MOVEMENTS

These like the segmenting movements are simple annular constrictions. They travel up and down short lengths of the bowel in a to and fro fashion at a rate of about 5 cm per second. They have the effect of carrying the chyme alternately from one to the other end of a loop of bowel. Their frequency is from 20 per minute in the duodenum to about 10 per minute in the lower ileum. The pendular movements are not always well marked. They are best seen in the rabbit, though they occur also in other animals and sometimes in man. At times they are very active and appear to throw the food material rapidly and with considerable force from end to end of a small section of the bowel. They constitute another factor contributing to the thorough mixing of the food, they have no direct effect upon the movement of food through the digestive tube, but exert a purely local churning or "shaker-like" action.

MOVEMENTS OF THE INTESTINAL VILLI

If the intestinal mucosa of a living animal be exposed and examined under the low power of the microscope the villi can be seen to be in constant motion—swaying from side to side, shortening and elongating alternately, or lashing to and fro, either singly or in groups. The movements are due ap-

parently to rhythmical contractions of smooth muscle fibers situated within the villi. By constantly stirring the fluids which bathe the surfaces of the villi, these movements undoubtedly aid the digestive and absorptive processes (see also pp 532 and 588). The movements are stimulated by contact with the chyme or with acid. Movements of the villi of an isolated, denervated loop occur when acid is placed in the duodenum, a hormone called *villikinin* has, therefore, been postulated but its existence has not been definitely established.

THE METABOLIC GRADIENT THEORY OF ALVAREZ

Alvarez believes that the fundamental factor determining the polarity of intestinal peristalsis is the gradual diminution in the metabolic activity of the intestinal muscle that occurs from the duodenum downwards. A parallelism was found between the magnitude of the energy exchanges of the bowel at different levels and its muscular activity. Rhythmicity, irritability and force of contraction and tone are graded from above downwards, being high in the duodenum and low in the ileum. The length of the latent period of the muscle also shows an increase as the intestinal tract is descended. The carbon dioxide production and oxygen consumption exhibited a corresponding decrease from the duodenum to the lower reaches of the intestine. The frequency of the rhythmical segmenting contractions (p 579) is 17 per minute in the duodenum but only 10 in the lower ileum. Warming a strip of ileum raises its metabolism, its rate of heat is increased to equal that of the duodenum while the latent period is shortened. The decline in the metabolic rate from duodenum to ileum is spoken of by Alvarez as the *metabolic gradient*, to it are ascribed the progressive diminution of the other physiological activities of the muscle. Upon the difference between the intensities of the latter at any two levels of the digestive tube, the direction and swiftness of the peristaltic wave are supposed to depend. Alvarez suggests in illustration of his conception, water flowing between two regions of unequal pressures. When the gradient has the normal slope, peristalsis is active and downward in direction. An inflammatory or irritative lesion, however, according to this theory, may raise the metabolic rate at a certain point so that it approaches, equals, or rises above that at a point nearer the pylorus. In such a circumstance the gradient would become reduced, annulled or even reversed. Sluggish peristalsis, stasis or antiperistalsis, respectively would result. Reversal of the gradient and the production of antiperistalsis as the result of mechanical blockage of the lumen, or injury and inflammation of a region of the bowel are advanced to explain the passage of fecal material into the stomach in intestinal obstruction and paralytic ileus.

The theory is interesting and suggestive. It accords

with many experimental and clinical observations, but cannot be said to have been established.

THE ILEO-COLIC VALVE (VALVULA COLI)

This structure permits the passage of the contents of the small intestine at intervals into the cecum, but when competent hinders the return of the material into the ileum. Its ability to do this depends, according to one view, not upon any mechanical valve-like device, but upon the contraction of the circular fibers of the gut which are thickened in this region to form a sphincter guard for the ileo-cecal orifice. But, though a sphincter-like thickening of the circular muscle fibers is found in this situation, most anatomists agree that the competency of the ileo-colic valve is mainly, or at least partly due to a valve-like construction. The ileum enters the cecum obliquely and in doing so invaginates the cecal wall, this alone would tend toward a valve-like action. The valve proper is formed as follows. As the lower end of the ileum enters the cecum, the invaginated portion of the cecal wall forms two transverse folds or lips, one above the other on the cecal aspect of the ileo-colic orifice. The lips fuse laterally to produce a fold on either side of the orifice (*the frenula coli*) which are continued round the interior of the cecum. As the cecum distends the frenula are stretched and pulling upon the lips from either side draw them firmly together. Thus the valve can withstand a high pressure in the cecum but yields to a low pressure from above. Materials introduced by enema may in some instances pass through the valve into the ileum. Such incompetence may permit the enema fluid to reach the duodenum. The ileo-colic opening has been observed in man through a cecal fistula. It appears as an oval or round opening from 2 to 3 cm. in diameter situated in the center of a small papilla. When tightly closed the valve was found to offer considerable resistance to the passage of the finger. While digestion was in progress the papilla was observed to flush, its color altering from a pale pink to a vivid red. The orifice opened rhythmically at frequent intervals and allowed a jet of fluid to escape into the cecum. Emotional excitement or the swallowing of food increased the frequency of the ejections (*gastro-ileal reflex*). During fasting nothing passed through for long periods, but in from $\frac{1}{2}$ to 4 minutes after food was taken into the mouth, fluid appeared in gushes of about 15 cc. every half minute or so.

The functions of the ileo-colic valve appear to be (a) to prevent the contents of the ileum from passing into the cecum before the digestive processes are complete (see also antiperistalsis in ileum, p 578), and, (b) to act as a barrier which prevents the bacteria-laden contents of the large bowel from contaminating the small intestine

THE MOVEMENTS OF THE LARGE INTESTINE

The contents of the ileum after passing through the ileo-colic orifice collect in the blind end of the cecum. In the human cecum little or no movement can be seen, as a rule, yet the material passes slowly into the ascending colon. This is due in part to filling of the cecum and passive overflowing but in many instances the material appears in the ascending colon *before* the former becomes filled. The forces bringing this about are not known. Todd describes slow changes in the shape of the cecum but thinks such a movement inadequate for the propulsion of its contents. Barclay remarks upon the absence of activity in man which, as he expresses it, presents a picture of "still life." In animals constricting rings may be seen passing over the cecum driving material against its blind end. These have the effect of churning the contents. Such antiperistalsis though not seen normally in the human cecum may, according to Todd, appear when the colon beyond is in a spastic state.

The rest of the human large intestine as revealed by radioscopy is usually free from antiperistaltic, segmenting or pendular movements, and peristalsis is absent or ill-defined. Some hours after a meal the shadow cast by the ascending, transverse and descending colons has usually a segmented appearance—a number of short sections strung together (fig 43 1). This appearance is caused by the haustra or sacculations in the bowel wall. Todd describes slow weak peristaltic movements and alternate shortenings and elongations—concertina-like action—in the transverse colon.

MASS PERISTALSIS

This is a movement first described by Holtz-knecht which occurs from time to time in the colon and sweeps the contents *en masse* for considerable distances along the bowel. The movement is analogous to the peristaltic rush already described as occurring in the small intestine. The mass movement in the colon of which the subject is quite unaware occurs only at fairly long intervals, probably not oftener than two or three times in twenty-



FIG 43 1 X-ray photograph of normal large bowel completely filled (After Barclay)

four hours. It is frequently preceded by the corresponding movement in the small intestine. It may occur almost immediately after the entrance of food into the stomach—*gastro-colic reflex*. The desire to stool so commonly experienced after breakfast is the result of this reflex; fecal material is forced into the rectum and the defecation reflex initiated (p 582). Mass peristalsis in the colon may also be induced by purely psychic influences—the taste, smell or thought of food or some disturbing emotion. Alvarez cites an instance of a subject with an incompetent anal sphincter in whom a bowel movement could be precipitated by the mention of savory food.

The movement usually starts in the region of the hepatic flexure. The haustral markings suddenly disappear, the bowel appearing radiologically as a solid unsegmented column. A strong and rapid peristaltic wave then travels over the transverse and descending colons carrying all before it. The haustral markings then reappear (fig 43 2). The contents of the more proximal portion of the colon are thus transferred to the pelvic colon which becomes filled from below upwards. The pelvic colon serves as a storehouse for the feces; it is evacuated through the rectum and anal canal by the act of defecation. The descending colon is usually empty except during the time that the feces are being transferred by a mass movement. Feces after they have filled the pelvic colon may, however, extend upward into the descending colon, and may even

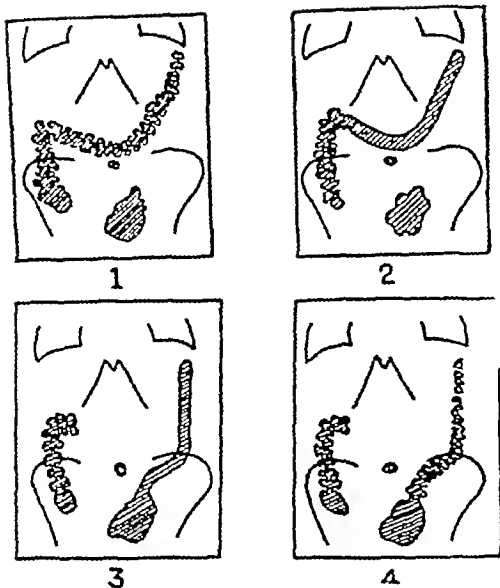


FIG. 432 Holzkecht's diagrams of the happenings during a mass movement of the large intestine 1, distribution of the food before any change was noted, 2, the haustral segmentation in the transverse colon has disappeared, 3, the whole colon beyond the hepatic flexure passes on suddenly, 4, it is again a picture of "still life" a few seconds later and the haustral segmentation has returned (After Barclay)

reach the splenic flexure. Except preceding and during the act of defecation the rectum also is normally empty.

DEFECATION

The act is brought about by the passage of fecal material past the resistance of the pelvi-rectal junction. According to some observers a definite thickening of the circular fibers of the colon (pelvi-rectal sphincter) exists in this situation. The entrance of the feces into the rectum may result from a mass movement already described or simply from overloading of the pelvic colon and the gradual pushing of its contents downward. The fecal masses fill and distend the rectum, and, when the intra rectal pressure reaches to between 40 and 50 mm Hg, the defecation reflex occurs. The reflex consists of a strong peristaltic contraction of the colon, accompanied by shortening of its longitudinal fibers which are disposed in three bands—*taeniae coli*—on its anterior, inner and posterior aspects. The *taeniae* obtain fixed points for their contraction indirectly from the coccyx, through the recto-coccygeus muscles. The latter arise from the coccygeal vertebrae and blend at the lower part of the rectum with the longitudinal fibers of bowel

which in this situation have become spread out to form a continuous sheathing for the rectum and anal canal. The contraction of the longitudinal fibers and consequent shortening of the bowel combined with the simultaneous downward progress of the peristaltic wave, and a coördinate relaxation of the anal sphincters, effects the evacuation of the feces. Hurst made observations upon human subjects by means of the X-rays and found that the evacuation of the colon extended as far or even beyond the splenic flexure. The movements of the bowel wall itself are accompanied and assisted by the contraction of voluntary muscles. The diaphragm descends to its extreme inspiratory position and is held there, the glottis being kept closed during the straining efforts which follow. The abdominal muscles contract powerfully and the intra-abdominal pressure rises. The descent of the diaphragm, as Hurst has shown, causes marked depression of the ascending and transverse colons and of the splenic flexure. The compression of the ascending colon against the nearly stationary cecum forces the former into an almost globular form. The pressure in the rectum, which in an individual in the erect posture amounts at ordinary times to about 20 mm of mercury, rises to from 100 to 200 mm during the straining efforts. The pelvic floor is supported against the elevated intra abdominal pressure by the contractions of the levatores ani, the transverse perinaei and coccygeus muscles. The levatores ani, which pass from their origins downwards on either side of the rectum and anal canal have some of their fibers inserted into the walls of these portions of the bowel or into the fibrous tissue in the neighborhood. By their contraction the anus is drawn upwards over the fecal mass. They also prevent prolapse of the rectum and anal canal, and by constricting the walls of the latter, at the termination of the act effect the expulsion of any material still remaining. The levatores ani serve also to reinforce the sphincters against the action of strong peristaltic waves and so enable the individual, when necessary, voluntarily to resist the evacuating reflex (see also p 585).

The rate of progress of a barium meal through the human intestinal tract

The time required for the first part of a barium meal (semi fluid) to pass from the stomach to the pelvic colon varies in different individuals. The usual time is from 12 to 14 hours. The material commences to leave the stomach almost immediately after it has been swal

lowed, it moves steadily and at a fairly rapid rate through the duodenum and very rapidly through the jejunum. Its progress through the ileum becomes progressively slower as the ileocecal opening is approached, and in the lower part of the ileum the material tends to accumulate before it is passed into the cecum. It commences to enter the latter in about $2\frac{1}{2}$ hours on the average. In 4 hours or so the material arrives at the hepatic flexure, and in about 6 hours at the splenic flexure.

Evacuation occurs at variable times after the material has commenced to collect in the pelvic colon but occurs usually in from 16 to 24 hours after the meal has entered the stomach. The first parts of the meal appear in the stools about 24 hours after the food has been ingested and the last traces in from 36 to 48 hours.

Many factors influence the rate of progress of food along the intestine—(a) Its chemical and physical characters (b) Individual variations in the activity of the intestinal musculature, or in the absorptive function of the colon, and consequently, in the consistency of its contents (c) The state of fullness of the alimentary tract. After a fast or after the bowels have been cleared by a cathartic the intestinal contents move more rapidly but the time elapsing before it is evacuated may appear to be longer than usual, since the pelvic colon takes time to refill (d) Muscular exercise increases the rate of movement (e) Emotional states also tend to increase intestinal activity.

THE ACTIONS OF CERTAIN SUBSTANCES UPON INTESTINAL MOTILITY. Fats, through the production of glycerine and soaps, stimulate the movements of the small bowel, sugars are also excitatory. Acetylcholine and pilocarpine, or neostigmine and mechoyl, are powerfully stimulant. Pituitrin is also excitatory whereas adrenaline is inhibitory. Ganglion-blocking agents, such as tetraethylammonium chloride, are powerfully inhibiting to all types of intestinal motility.

Atropine reduces the tone of the musculature of the small and large bowel and relieves intestinal spasm. The phasic contractions of the small bowel are also usually decreased in amplitude but sometimes appear to be increased, in any event the propulsive force of the contractions is diminished, so that the progress of material along the intestinal lumen is slowed. The movements of the large bowel, especially of the distal colon, are depressed. Morphine increases intestinal tone, the rhythmical movements (segmenting and pendular) are increased in frequency, but reduced in amplitude. The propulsive movements of the intestinal tract are profoundly depressed or abolished, constipation results.

The stimulating effects of laxatives upon the bowel movements are well known, but their mode of action is various, and in some instances obscure. *Epsom Salts* (MgSO_4), *Sodium sulfate* (Na_2SO_4) or *phosphate* (Na_2HPO_4) and other saline cathartics which undergo but slight absorption add water to the intestinal con-

tents (and consequently increase their bulk) through the change in osmotic relationship between the blood and intestinal fluids. *Agar, bran, fruits and vegetables* which leave indigestible residues—seeds, skins, fibrous material, etc.—also stimulate the bowel movements by increasing the bulk of the feces. *Oils*, especially mineral oil which is undigested and virtually unabsorbable, act mainly as lubricants, and to prevent undue absorption of water from the feces in the large intestine, and thus to avoid their becoming dry and hard. Castor oil contains an irritant principle (ricinoleic acid) to which its stimulating effect upon the bowel is due, this principle is released when the oil is digested by the intestinal lipases. Castor oil is but one of a large class of cathartics which depend for their action on an irritant principle. It includes aloes, senna, rhubarb, cascara, jalop, phenolphthalein, and a number of others. Many such drugs through their stimulating effect upon the mucous secreting cells of the intestinal lining also increase the water of the intestinal contents.

THE NEURO-MUSCULAR MECHANISMS GOVERNING THE MOVEMENTS OF THE INTESTINAL TRACT

INNERVATION OF THE SMALL INTESTINE

The wall of the small bowel is composed of an outer longitudinal and an inner circular coat. The latter is considerably thicker than the former. According to Carey both coats are arranged in a spiral fashion. Between the two muscular sheets the *myenteric plexus* (*Auerbach's*) is situated. The *plexus of Meissner* lies in the submucosa. These two plexuses are connected with one another by nerve filaments which pass between the circular muscular fibers. Auerbach's plexus contains numerous ganglion cells, which are scarce in Meissner's plexus.

The rhythmical contractions—*segmenting* and *pendular*—are myogenic, that is, they are dependent solely upon the rhythmical property of the intestinal muscle itself. They are not abolished by such nerve poisons as cocaine and nicotine. A segment of excised intestine beats rhythmically when immersed in a 1 to 4000 cocaine solution. Furthermore, the contractions of the circular coat of the bowel have been shown by Gunn and Underhill and by Alvarez and Mahoney to continue after it has been stripped from the longitudinal layer, and from the submucosa as well, all ganglion cells are in this way removed (See also Gasser, and Thomas and Kuntz).

The peristaltic contractions, on the contrary, are dependent upon the intrinsic nerve plexuses



FIG 43.3 Upper, shows the inhibitory effect of splanchnic stimulation upon the movements of the small intestine, the heavy white line indicates the time during which the stimulus was applied (after Starling) Lower, shows effect upon intestinal motility of stimulating vagus nerves (After Thomas and Kuntz)

But though carried out through local reflexes in the bowel wall (myenteric reflexes of Cannon) they are readily influenced through the extrinsic nerves—the *vagus* and *sympathetic*. It was shown by Bayliss and Starling that section of both these sets of nerves does not abolish the peristaltic movements, whereas the application of cocaine to the bowel wall does. The *vagus*, whose terminals connect with ganglion cells in Auerbach's plexus, augments the movements. The sympathetic is inhibitory (fig 43.3). The sympathetic fibers pass mainly in the lesser splanchnic nerve to the lower part of the celiac ganglion and to the superior mesenteric ganglion (fig 42.7), from these ganglia the impulses are relayed to the bowel. The post-ganglionic fibers pass without interruption through the plexuses of the bowel wall to end in direct relation with the muscle cells. The extrinsic nerves also influence the tone of the intestine, the *vagus* increasing, the sympathetic diminishing this property. The sympathetic exerts a continuous inhibitory action upon the bowel movements which are therefore augmented after section of the sympathetic fibers. It has also been shown by Campbell

and Markowitz that spinal anesthesia which temporarily paralyzes the thoraco lumbar outflow removes the reflex inhibition of the intestine induced in dogs by peritoneal irritation (intra-peritoneal injection of tincture of iodine). The *vagus* on the other hand does not appear to exert any continuous augmentor effect upon intestinal motility, for section of these nerves does not affect the movements.

INNERVATION OF THE COLON

In the cecum and the rest of the colon the fibers of the outer muscular coat are gathered into three longitudinal bands—the *taeiae coli*—which, being shorter than the underlying layer, draw the bowel into the tucks and pockets known as haustra. The circular coat is thickened between the latter. The intrinsic nerves (plexuses) of the colon have a distribution similar to that of the small intestine.

The extrinsic nerves

For a variable distance from its commencement the large intestine is supplied with motor fibers through the *vagus*. The vagal innervation terminates usually within the first half or third of the transverse colon. The rest of the colon, including the rectum, receives its motor innervation through the pelvic nerves (*nervi erigentes*, ch. 72) from the 2nd, 3rd and 4th sacral segments. The pelvic nerves are cholinergic.

Inhibitory fibers to the entire colon are derived from the *sympathetic*. They arise from the lumbar segments of the cord and reach the proximal part of the colon (cecum, ascending and transverse colons) through the inferior mesenteric plexus. The fibers to the distal colon (the descending, iliac and pelvic colons and the rectum) arise from the 2nd and 3rd lumbar segments. They pass via the lumbar splanchnics to the inferior mesenteric ganglion and thence to the bowel by (a) a number of short strands called the *lumbar-colonic nerves* (also known as the *inferior mesenteric nerves*) and (b) the *hypogastric (presacral) nerve*. The lumbar-colonic nerves are the axons of cells situated in the inferior mesenteric ganglion.¹ The fibers entering into the formation of the hypogastric nerve pass for the most part without interruption through the latter ganglion. Their cell stations are situated in the hypogastric ganglion. The postganglionic

¹ According to Learmonth there can rarely be found, in man, a definite structure to which this name can be applied.

fibers pass through the pelvic plexus to the bowel (figs 42 7, 43 4) The hypogastric nerves, and probably the lumbar colonic nerves as well, are adrenergic

Stimulation of the lumbar-colonic nerves causes relaxation of the distal colon Learmonth and Markowitz showed that these nerves exert a constant inhibitory action since increased colonic activity follows their section Garrey found that though in the decerebrate cat the distal colon is inactive when its sympathetic supply is intact, division of the lumbar-colonic nerves caused rhythmical activity and a marked increase in tone The hypogastric nerve appears to exert a minor inhibitory influence upon the colon since its division causes only a slight increase in the activity of the bowel Paralyzing the lumbar outflow by a spinal anesthetic has an effect similar to that of nerve section² The inhibitory impulses of the colon arise apparently within the lumbar cord, for if this region has been isolated previously from higher centers by spinal transection the full augmentor effect upon the colon of sectioning the colonic nerves is obtained

Section of the *pelvic nerves* (sympathetic innervation intact) relaxes the wall of the distal colon and the animal subsequently experiences difficulty in emptying the bowel The effect of section of the pelvic nerves upon the tone of the colon is particularly well shown if the bowel has been previously in a hypertonic state as a result of division of the sympathetic supply Section of the cord above the sacral segments also causes relaxation which indicates that the constant augmentor effect is due to impulses arising in higher centers A subsidiary augmentor center apparently exists, however, in the sacral cord, for cutting the pelvic nerves some time after spinal transection causes colonic relaxation That is, the sacral segments acquired control during the time which had elapsed after their isolation from higher centers

INNERVATION OF THE SPHINCTERS

The sphincters of the bowel—ileo-colic, and the internal and external anal sphincters—exemplify the principle of reciprocal innervation, or what Meltzer termed when speaking of muscular tubes, "contrary innervation" For instance, the motor nerve to the *ileo-colic* sphincter is the same as that

² It is likely that the center for the sacral outflow is also paralyzed by the anesthetic but that the bowel movements are then controlled by the intrinsic plexuses

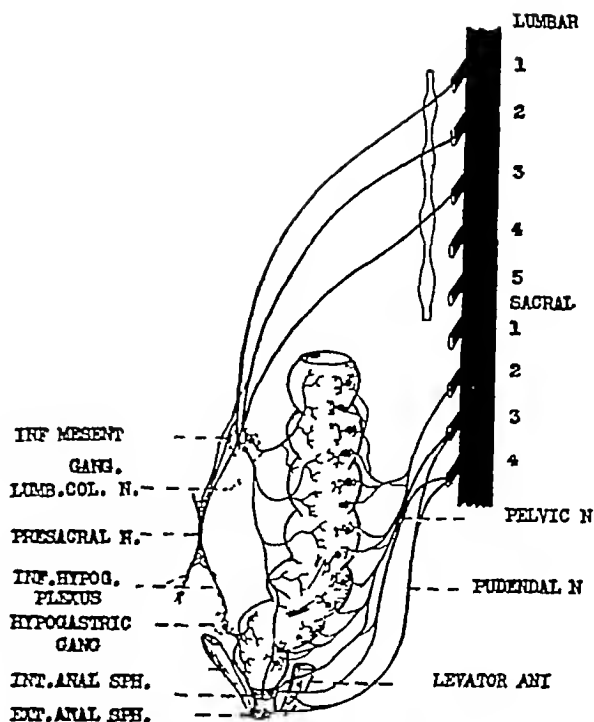


FIG 43 4 Diagram of the innervation of the distal colon

which causes inhibition of the bowel wall, namely, the sympathetic Stimulation of this nerve will in consequence produce a dual effect, contraction of the sphincter and inhibition of the wall of the ileum In this way the passage of fluid into the cecum is restrained Logically we should expect the vagus to have an inhibitory effect upon the ileo-colic sphincter, but this action has never been demonstrated, and so far as is known the vagus is without any effect upon the sphincter

The *internal anal sphincter* is also innervated in a fashion contrary to that of the rest of the large bowel The pelvic is the inhibitory nerve, while the sympathetic is motor The sympathetic fibers are conveyed via the lumbar-colonic and hypogastric nerves Stimulation of parasympathetic fibers will therefore induce evacuation of the bowel by excitation of the bowel wall and relaxation of the sphincter Stimulation of the sympathetics on the other hand inhibit intestinal activity and increase the tone of the sphincter, the act of defecation will be restrained The *external anal sphincter* is composed of striated muscle, it is kept tonically contracted and is under voluntary control It is said not to atrophy when its nerve supply (derived through the 4th sacral nerve and the perineal branch of the pudendal nerve) is cut, but to recover its tone to a large extent after the operation Its

contraction, with that of the levator ani and other perineal muscles, is an important factor in opposing defecation. Afferent impulses arising in these muscles during their contraction are said to reflexly inhibit the bowel movements.

THE INNERVATION OF THE MOVEMENT OF DEFECATION

The defecation reflex is governed by a medullary center and a subsidiary center in the cord. The existence of the latter is evident in nervous diseases associated with complete functional or anatomical division of the cord above the lumbosacral region, when evacuation is effected automatically and in an almost normal manner. Also in animals after complete transection of the cord, though there is loss of control over the bowel movements and incontinence of feces for a time, the spinal center subsequently assumes control and automatically evacuates the bowel at regular intervals.

The *spinal center* is situated in the second, third and fourth sacral segments, but the destruction of this region in animals is not necessarily followed by complete and permanent fecal incontinence, for the rectum like the bladder (p. 485) is capable of controlling evacuation through its intrinsic nervous mechanisms. Also in involvement of the sacral segments of cord in man there sometimes may be adequate automatic control of the rectum and anal sphincter.

The *medullary center* is situated in the floor of the fourth ventricle not far from the vomiting and respiratory centers. Hatcher has shown that a stimulus to defecation will inhibit vomiting and certain emetic drugs (picrotoxin) applied to the vomiting center cause defecation. Hatcher and Weiss found areas in the medulla which controlled the tone of the sphincters. The well known fact that stimulation of the anal sphincter by forcible dilatation is often an effective means of stimulating respiration, suggests a close relationship between the respiratory and defecation centers.

The reflex as mentioned on page 582 is initiated under normal circumstances by the passage of feces into the rectum. The latter, however, soon adjusts its capacity to the bulk of the feces (postural tone) the pressure stimulus being then abolished. The rectum may therefore become unresponsive¹ if defecation is voluntarily prevented.

¹ This was well shown in the human subject by Hurst. When the rectum was inflated the call to defecate became urgent, but the intrarectal pressure fell again within half a minute or so, as a result of the adaptation

The afferent fibers conveying impulses from the rectum travel mainly if not entirely with the efferents of the pelvic nerve and enter the cord by the posterior roots of the sacral nerves. Afferent fibers supplying the anal canal probably travel in the pudendal nerves. Division of the main afferent paths abolishes the defecation reflex since no impulses can reach the center. There is retention of feces for a time, but as mentioned above the local nervous mechanisms of the bowel wall may later assume control.

The reaction of the intestinal contents. The reaction of the contents of the duodenum vary considerably in accordance with the acidity of the chyme entering from the stomach and with the volume of the alkaline fluids—especially of the pancreatic juice. The proportion of acid and alkaline fluids varies in turn with the stage of digestion and the type of food, meat tending to increase the acidity of the chyme. In the dog the pH of the duodenal contents may be anywhere from 4.6 to 7.6, the usual value is about 6.⁴ In the human subject, the type of food and differences in the secretory activity of the gastric glands cause considerable individual variations in the reaction of the duodenal contents, in subjects of subacidity or hyperacidity (p. 519), for example, the reaction tends to be more alkaline or more acid, respectively. Compared with normal persons, in patients with duodenal ulcer the pH of the duodenal contents tends to be lower and to remain at the lower level for a longer time. Samples of human duodenal contents, obtained by duodenal tube, show in most cases a pH between 5 and 6, average about 5.5, but may reach a pH of 2 for short periods. Determinations upon healthy animals (rats and dogs) indicate that the contents of the entire small intestine below the duodenum are almost always definitely acid in reaction. From observations made upon patients with intestinal fistulae, this is most probably true also of the human intestine. The acidity of the small intestine (except in its upper part) is due to organic acids resulting from the action of fermentative types of bacteria (p. 593). In rats there occurs a progressive rise in pH from the upper to the lower levels of the small intestine. The average pH of the contents of the former is around 6 and of the latter about 6.8. The pH falls slightly in the cecum but rises again in the colon. The feces have a slightly acid, a neutral or a faintly alkaline reaction (6.9 to 7.2). In rachitic animals the pH of the intestinal tract and of the feces is higher than that of normals. The administration of cod liver oil or exposure to sun

of the rectal walls to the distending force, and the sensation passed off. Rapid re-inflation caused the sensation to return and then pass away again as the pressure fell.

⁴ It will be noted that this is not the optimum pH for the action of the pancreatic enzymes.

shine restores the reaction to the normal level (Zucker and Matzner)

ABSORPTION FROM THE GASTRO-INTESTINAL TRACT

Ethyl and methyl alcohols are absorbed at a fairly rapid rate from the human stomach. The gastric absorption of water is considerable. In dogs, half of the amount of *heavy water* (D_2O) administered is absorbed within 20 minutes (Cope and associates). Ordinary water, no doubt, is absorbed to the same extent. The gastric absorption of *glucose* and other sugars is small, but by no means negligible. The absorption increases with the concentration of the solution used, being inappreciable in the case of glucose below a concentration of 10 per cent. Of the sugars investigated by Feitelberg galactose was most rapidly absorbed (about 24 per cent of 22 cc of a 21.5 per cent solution in an hour). Next in order of absorbability came glucose, lactose, levulose and sucrose. In no instance was the absorption sufficiently great to raise the blood sugar.

In the study of gastric absorption the stomach is ligated at the pyloric and cardiac ends and the material to be investigated introduced through the gastric wall by means of a syringe. The extent of the absorption is determined by ascertaining the quantity of the material which has disappeared within a stated time. The result obtained does not therefore indicate the quantity of the material which would be absorbed under normal conditions. A solution of glucose, for example, if isotonic, passes rapidly into the duodenum of the intact animal and if hypertonic is first rendered isotonic by the gastric secretion and before being evacuated. The results of different investigators upon the absorption of *hydrochloric acid* from the stomach are conflicting but it seems likely that a considerable reabsorption of this acid does occur. *Hydrocyanic acid* is rapidly absorbed in fatal amounts. *Strychnine* is also absorbed from the stomach in sufficient amounts to cause death within a few minutes.

Fats even in fine emulsion, e.g., egg-yolk, cream, and peanut oil, are not absorbed from the stomach into the circulation, although some attempt at absorption occurs, for fine droplets of the administered fat appear in the gastric mucosa and submucosa. Unchanged *proteins*, e.g., egg-albumin, pass through the gastric mucosa but usually in such minute amounts that delicate immunological tests, e.g., precipitin reaction, with the serum of animals sensitized to the particular protein, are required to detect their presence. Split products of

protein, e.g., proteoses, peptones, peptides and amino-acids, are also absorbed from the stomach in insignificant amounts.

It may be concluded then that with the possible exception of sugars, gastric absorption of the main types of food during normal digestion is negligible. The mucosa of the small intestine is the only important portal through which food materials enter the body.⁵ Absorption from the colon is restricted mainly to water and inorganic salts. Glucose solution (10 per cent) introduced by enema may be absorbed in considerable amounts but the absorption of proteins and fats is for practical purposes (rectal feeding) unimportant. The absorption of the different food materials from the small intestine is considered in the respective sections dealing with their metabolism.

The villi are the absorbing units of the small intestine. Each of these minute finger-like processes is covered by a layer of columnar epithelium, and contains a capillary loop and a lymph vessel (lacteal), a small amount of connective tissue and strands of smooth muscle. There are some 5,000,000 villi in the human intestine which provide a total absorbing surface of some 10 square meters (fig 401, p. 533). Amino-acids and glucose are absorbed into the capillaries, about 60 per cent of the digested fat can be recovered from the thoracic duct, the remainder is absorbed apparently into the portal blood.

Attempts have been made to explain intestinal absorption by known physico-chemical laws. The factors of a physico-chemical nature whereby absorption might conceivably be brought about are (a) Differences in concentrations and therefore of the diffusion pressures between crystalloids in the blood and in the intestinal lumen. (b) The hydrostatic pressure within the intestine created by the intestinal movements, which would tend to drive water and dissolved substances across the epithelial boundary. Hamburger found that absorption varied directly with the intra-intestinal pressure, and Wells has shown that the absorption of water from isotonic sodium chloride solutions ceased when the intra-intestinal pressure was reduced to from 8 to 26 cm. below that of the atmos-

⁵ The enormous reserve of the digestive and absorptive capacities of the small intestine has been demonstrated in a striking fashion. After some 19 feet of small intestine (total length about 22 feet) together with the cecum, ascending colon and part of the transverse colon were removed from a patient for malignant disease, there resulted little disturbance in absorption and no loss of weight.

phere, but that the absorption rate increased in proportion to the increase in intra-intestinal pressure above this level (c) The osmotic pressure of the plasma proteins in excess of the hydrostatic pressure of the capillary blood pressure was suggested by Starling as an important factor in attracting water and crystalloids into the blood stream It has been estimated that the colloid osmotic pressure is some two or three times greater than the blood pressure in the capillaries of the villus Wells found that dilatation of the mesenteric vessels reduced absorption, and attributed the reduction to an increase in the hydrostatic pressure of the capillary blood On the other hand, Van Stickney and associates observed an increased absorption of saline when the general blood pressure was raised to between 150 and 190 mm Hg, which they thought was probably the result of the opening up of additional capillaries and the consequent expansion of the absorptive surface (d) Hober suggests that electrical forces in the nature of cataphoresis (the transfer of a solute across a membrane from anode to cathode) play an important rôle

Though some or all of the factors enumerated above probably play a part in the absorption process, there are several observations which cannot be explained upon a purely physico chemical basis, some specific "vital" activity must be attributed to the epithelial cells of the villi For example, materials such as glucose and sodium chloride are absorbed from hypotonic solutions (i.e., solutions containing the material in lower concentration than it exists in the plasma) Also, as shown by Cori, the absorption of glucose from the intestine may continue at an almost constant rate for hours though its concentration diminishes progressively, yet the absorption of a material, such as glucose, can not be increased by simply reducing its concentration in the plasma Such facts are not explained by the laws of diffusion Moreover, Hewitt has shown that when glucose, galactose and fructose in equivalent concentrations are placed in an intestinal loop the relative absorption rates are glucose > galactose > fructose After destruction of the epithelial lining all three sugars are absorbed at the same rate Macleod and associates found that glucose was absorbed more rapidly from a surviving intestinal segment than the foreign sugars xylose, mannose and arabinose This preferential absorption of glucose disappeared when the temperature of the

segment was lowered to 0°C The sugars then passed through the intestinal wall in accordance with their diffusion rates through a dead membrane—xylose with the smaller molecule diffusing more rapidly than glucose

That the epithelial covering of the villus is not merely a passive membrane, but performs a specific function which entails the expenditure of energy, is also indicated by the experiment of Brodie, Cullis and Halliburton They observed an increase of over 150 per cent in blood flow, oxygen consumption and carbon dioxide production during the absorption of salt solutions of different concentrations They concluded that a specific function of the epithelial lining, rather than contractions of the wall of the intestine was responsible for the increased metabolism In the more recent experiments of Van Liere and his colleagues, it was found that oxygen lack does not impair the absorption of all substances alike Anoxic anoxia to a degree compatible with life does not reduce the absorption of glucose and actually increases the absorption of glycine. Nor is the absorption of sodium sulphate reduced by anoxia A phosphorylation process rather than an oxidative one is probably concerned in the absorption of glucose, which would account for its being unaffected by anoxia Anoxic anoxia definitely depresses the absorption of sodium chloride, but anoxia produced by hemorrhage, on the contrary increases absorption of isotonic salt solution

Any consideration of intestinal absorption recalls the process of selective reabsorption by the cells of the renal tubules and suggests that similar factors are concerned That the cells of the intestinal mucosa perform chemical work, as in the absorption of fat (p 693) is accepted Reabsorption of water and dissolved substances from the glomerular filtrate is, one might say, a process of reverse secretion—from tubular lumen to blood The laws of diffusion are just as inadequate to account for the absorption from the intestinal lumen as they are to explain the reabsorption process in the kidney

THE FORMATION OF FECES

The contents of the ileum have an almost liquid consistency The quantity of fluid material passing through the ileo colic valve in 24 hours is in the neighborhood of 400 grams The feces evacuated during an equivalent length of time is about 150 grams The fluid is absorbed mainly by the cecum

and ascending colon, but also in smaller amounts during the progress of the feces along the transverse and descending colons and during their stay in the pelvic colon. The feces are neutral, slightly acid or slightly alkaline in reaction (p 586)

The volume of gas contained in the human gastro-intestinal tract under ordinary conditions has been estimated at about 1000 cc. It is composed of carbon dioxide (23 per cent), nitrogen (21 per cent), oxygen, hydrogen and methane.

The feces are composed of food residues, bacteria, material secreted through the wall of the intestine and in the bile, leukocytes and desquamated epithelial cells. Food residues constitute a much smaller proportion of the bulk of the feces than is usually realized. The fat, protein and carbohydrate of the diet is practically all absorbed and if the food be free from indigestible material, especially cellulose, the feces are composed almost entirely of bacteria, secretions, etc. During starvation, for example, feces continue to be formed and their composition does not differ materially from that of feces passed after an ample diet. Also, a segment of a bowel when isolated from the rest of the intestinal tract becomes after a time packed with a mass of pasty fecal material, which of course must be entirely endogenous. Wide variations in the composition of the diet, if the quantity of cellulose remains unaltered, exert little or no influence upon the composition of the feces. The bulk of the feces is reduced however, during starvation, but the reduction is due largely to the removal of the stimulating effect of the food upon the secretory activity of the intestine. Indigestible materials, especially cellulose, increase the amount of the feces, not only by adding directly to their bulk, but also through increasing the amount of endogenous material.

Bacteria constitute ordinarily about 9 per cent of the total solids of the feces. The nitrogen of human feces amounts daily to between 0.5 and 1 gram and, as just mentioned, is largely endogenous. Fecal "fat" is also largely endogenous, continuing to appear in the feces though all fatty material has been excluded from the diet, it differs chemically from ordinary food fat but resembles closely the blood lipids. It consists mainly of lecithin and coprosterol, the latter being derived from cholesterol through the reducing action of bacteria. A certain proportion of the fecal fat is also derived through bacterial action from non-fatty sources. In human feces a part (cholesterol

and lecithin) is of biliary origin. Calcium, phosphate, magnesium and other inorganic materials in the feces are also derived mainly from the blood. Fecal iron is mostly exogenous (see p 76).

The feces owe their color chiefly to *stercobilin* (urobilin, p 539) produced by the bacterial reduction of bile pigment, their odor is due to the presence of aromatic substances, chiefly *indole* and *skatole*, derived through the bacterial deamination and decarboxylation in the large intestine, of *tryptophane*, but also to the production of mercaptan and hydrogen sulfide through the bacterial decomposition of cystine (p 590).

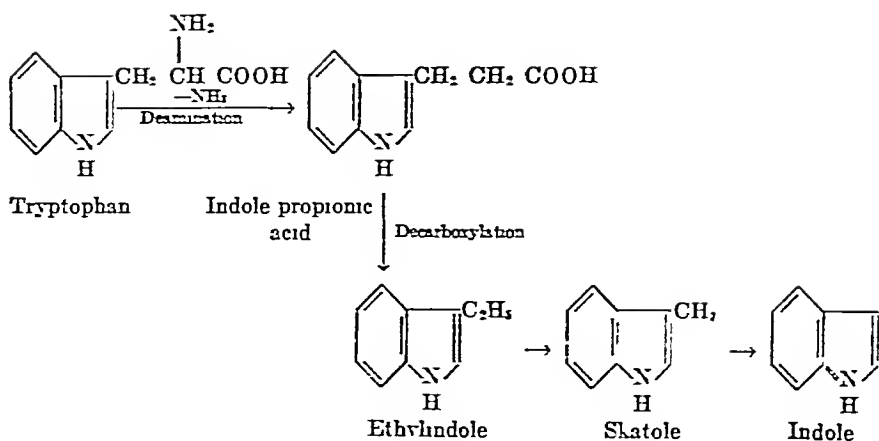
Small quantities of indole, skatole and phenol are absorbed into the blood stream but undergo detoxication in the liver by conjugation with sulphuric acid and potassium or with glycuronic acid. Phenol may undergo conjugation in either of these ways, in the former instance phenol sulphate of potassium is formed, in the latter the corresponding glycuronate results. Indole and skatole after absorption are converted to indoxyl which then for the most part combines with sulphuric acid and potassium to form *indoxyl sulphate of potassium* or *indican*, which is excreted in the urine. Indole and skatole are also conjugated but in relatively small amounts with glycuronic acid forming indoxyl glycuronic acid and skatoxyl glycuronic acids, respectively (See next page.)

Other toxic materials (amines) produced through the decarboxylation (without deamination) by bacteria of amino acids, are ethylamine from alanine, tyramine from tyrosine, histamine from histidine, cadaverine from lysine and putrescine from arginine.

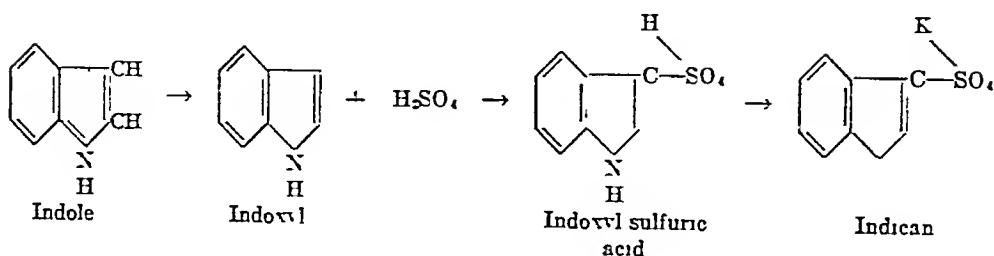
CONSTIPATION

One or more of many factors may be responsible for, or contribute toward the production of constipation. Apart from some gross mechanical obstruction, e.g., tumor, adhesive bands, strictures somewhere along the bowel, the chief causes are the following:

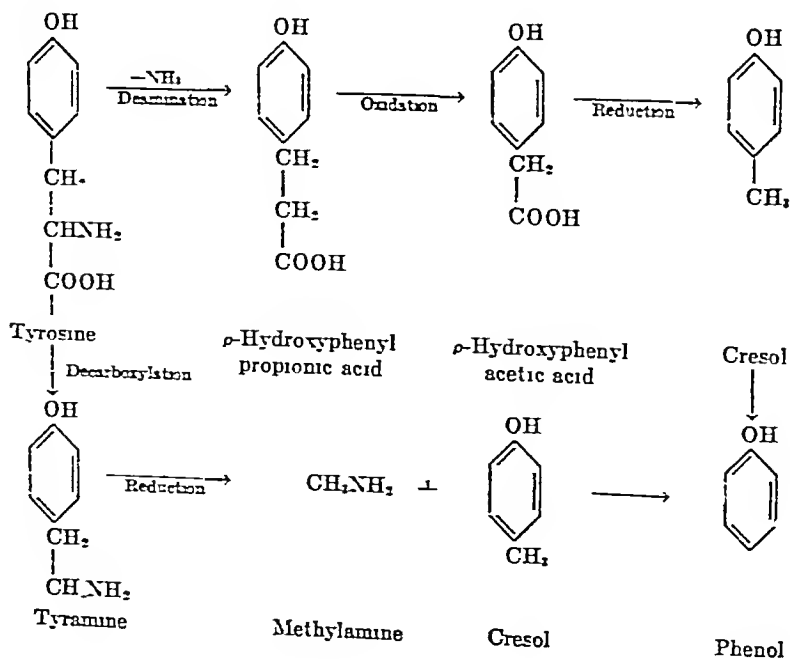
(1) *Habit*. The act of defecation can usually be readily inhibited by an effort of the will. The refusal to respond to the sensation aroused by a full rectum and the failure to acquire the habit of clearing the bowels regularly at some definite time each day, e.g., after the first meal, is a common cause of constipation. The rectum becomes tolerant, adapting itself to the increased fecal bulk (see footnote, p 586), and the desire to defecate



The formation of indican



Conversion of tyrosine to tyramine, cresol and phenol



passes. The retention of feces in the distal colon and rectum leads to the excessive absorption of fluid, the feces therefore become dry and hard and less easily expelled. There may at first be no defect in the musculature or movements of the digestive tract as a whole but the lower segment of the colon as a result of continued overloading becomes sluggish. Atony and thinning of the muscular wall may eventually result. To this type of constipation Hurst has applied *dyschezia* (meaning a difficulty in easing oneself). It is obvious then that properly regulated habits are more likely to correct this condition than the use of purgatives which stimulate peristalsis in an intestinal wall that is, in general, in no need of urging. By purgatives "the whole intestinal tract is teased and pained for the defective action of that part of it which is most remote from their influence."⁶

The dyschezia may be aggravated by other factors. Sometimes the angle at the junction of the pelvic colon and rectum is abnormally acute. On the other hand, spasm (or achalasia) of the circular fibers (sphincter) in this situation may be present which offers considerable obstruction to the passage of feces.

(2) A diet that leaves too little of the necessary residue for the stimulation of intestinal activity or one which contains too little fluid.

(3) A colon which absorbs too readily—the "greedy" or "thrifty" colon as it has been called—and causes undue condensation of the intestinal contents.

(4) *Hypertonic state of the musculature of the colon* is a cause of what is termed *spastic constipation*. Both the transverse and descending colons are sometimes held firmly contracted to form a solid cord which may sometimes be felt through the abdominal wall. At other times only the pelvic and descending colons are affected or again the spasm may be confined to the transverse colon. The contracted bowel is the seat of purposeless spasmodic contractions which have little effect in moving the food onward. The spasticity may be reduced by atropine which acts by depressing the parasympathetic influence upon the large intestine. The hypertonic state supposedly may be induced reflexly through impulses arising in diseased pelvic or abdominal viscera. On the other hand, it may be due to influences of a mental nature, overwork, worry or shock, etc. In the constipation of chronic lead poisoning a spastic state of the colon exists.

⁶ Chevalier, 1819, quoted by Hurst.

(5) *Weakness or depressed activity of the musculature of the colon—atonic constipation*. An inherent defect in the neuromuscular mechanism of the bowel is uncommon. Atony of the intestinal musculature may, however, be associated with certain general conditions, e.g., senility, obesity, etc., constitutional diseases or lesions of the central nervous system. The driving force of the intestinal muscle under such circumstances is subnormal. The delay in the movement of the food may be in the small as well as in the large intestine. A hypotonic state of the intestinal musculature may also result from a diet low in vitamin B₁ (p. 750). Rats fed on diets deficient in inorganic salts show loss of tone and depressed motility of the intestinal tract (Robertson and Doyle). The condition is corrected by adding Ca and K to the diet, which suggests that a low intake of these minerals may be a factor in some cases of chronic constipation in the human subject.

The subjective manifestations of constipation are considered elsewhere (see p. 592 and 1053).

CONGENITAL MEGACOLON—HIRSCHSPRUNG'S DISEASE. This is a condition appearing in childhood. It is characterized by an enormous dilatation of the colon and hypertrophy of its walls, with obstinate constipation. There is a congenital absence of the ganglion cells of Auerbach's plexus—the postganglionic neurons of the pelvic nerve. This defect involves the internal anal sphincter, the rectum and rectosigmoid, and a variable part of the more proximal colon; it may extend upwards to include part of the vagal innervation of the colon. The internal anal sphincter and rectum are contracted into a firm cord. Excision of this part of the bowel is advocated for the relief of the condition (see Swensen).

INTESTINAL INTOXICATION

THE LARGE INTESTINE. Several substances, some intensely toxic in character, are formed in the large intestine as a result of the decomposition of protein by the normal bacterial flora (colon type of organism). Among such putrefactive products are histamine, phenol, cresol, indole, skatole, ethylamine, isoethylamine, tyramine, etc. (p. 590). Choline is also formed as a result of the decomposition of lecithin, and choline gives rise to traces of neurine. Some observers, impressed by the powerful actions of these substances when injected into animals, have suggested that their absorption into the general blood stream is responsible for many ills, and particularly for the well-known symptoms of constipation. Metchnikoff arraigned the colon as

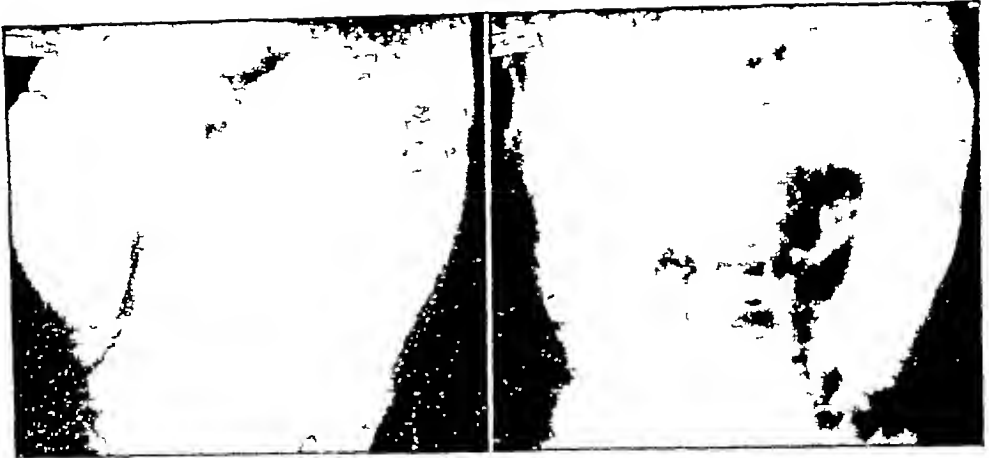


FIG 435 Megacolon. Motor test with spinal anesthesia. Left, the injection of three and one half quarts of fluid causes painless distension of abdomen. In horizontal position no spontaneous expulsion after fifteen minutes of effort. Right, fifteen minutes after spinal anesthesia. Note the emptying of the lower sigmoid and most of the ascending colon, also the reduction in diameter of the upper sigmoid (After Scott and Morton.)

the body's greatest enemy Arbutnot Lane in more recent times inveighed against this cess-pit and recommended its extirpation under certain circumstances. No reliable evidence can, however, be cited to connect any of the toxic products mentioned with the symptoms of disease. Many of them are produced only in mere traces, and even though some, such as indole, are produced in relatively large amounts (60 mg, according to Herter, may be present in 100 grams of feces) toxic doses are prevented from entering the systemic circulation by the barrier offered by the bowel wall and through the detoxicating action of the liver (p 589). Indole, for example, when given to the human subject by mouth in an amount (1 gram), which is more than would ever be present in the bowel at one time, produces no symptoms, 2 grams causes only some dizziness and slight headache, while large amounts may be introduced into the colon without ill effect. The slight absorption of indole and skatole under ordinary circumstances is indicated by the fact that though the feces contain considerable amounts of both these substances, the detoxicated product indican may be absent from the urine. On the other hand the ability of the liver to protect the body against these substances is shown by the observation that large amounts of indican are constantly present in the urine of certain individuals, showing that considerable quantities of the putrefactive products are absorbed, yet such persons enjoy good health.

Choline may be given by mouth with impunity,

even when relatively large amounts (15 to 20 mg per kg in the cat) are given intravenously; no toxic effects are produced. Since this substance is rapidly destroyed in the blood stream, large amounts may be injected slowly over a period of time without noticeable effect.

Histamine, another powerful depressor substance was found by Meakins and Harrington in minute amounts in the cecum, but only insignificant amounts, they conclude, pass through the intestinal wall. Even after the introduction of a large dose into the small bowel only slight absorption was evident. Hanke and Koessler gave the amine by mouth in doses of 100 mg to guinea pigs, there were no ill effects, and upon analysis of the intestinal contents 24 hours after the administration of this amount, less than 2 mg were recovered. The chief protection against the toxic effects of histamine is apparently in the bowel wall itself. It has also been found by Weiss and associates that in the case of the human subject 20 mg of histamine may be injected per hour without causing a depressor effect, 500 mg given orally were inert.

According to Alvarez the symptoms of constipation—headache, furred tongue, foul breath, malaise, etc., are of reflex rather than of toxic origin. Afferent impulses set up from the wall of the overloaded rectum appear to be responsible. Distension of the rectum by inert material such as absorbent cotton has been shown to produce almost all the symptoms of constipation. In the dog, packing the rectum in this way causes a rise

of some 10 mm Hg in blood pressure which is maintained until the foreign material is removed

The manifestations of constipation referable to the alimentary tract, namely, the loss of appetite, coated tongue and offensive breath are, according to Alvarez, probably due to small waves of contraction originating in the wall of the loaded bowel and traveling in a reverse direction over the small intestine, stomach and esophagus (p 561) A strong argument against the symptoms of constipation being due to toxic absorption is the almost immediate relief which follows evacuation of the bowels—it would certainly take some time for the blood to be cleared of a poisonous material Furthermore, in constipation the feces are relatively dry and hard, conditions much less, rather than more favorable to the absorption of putrefactive products One should expect, indeed, that the latter would occur more readily during diarrhea Apart from all the indirect evidence against a toxic element being concerned in constipation, is the positive fact that many subjects of the disorder, even though this is of a severe grade, do not excrete indican in the urine and, as mentioned above, other individuals who for some unknown reason excrete large quantities remain free from symptoms

There seems little doubt that toxic products of the bacterial flora of the large intestine are prevented from entering the systemic circulation in amounts that are pathologically significant

THE SMALL INTESTINE The immunity of the body to autointoxication applies only to the large intestine The small intestine is not equipped in the same degree to resist the passage of toxic products into the blood stream Normally, however, the bacterial flora of the small intestine are quite different from those of the large The microorganisms (e g, *Bacillus bifidus*) in the former situation have a fermentative not a putrefactive action Through their action upon carbohydrate, organic acids, acetic, butyric and lactic, are produced The acid reaction of the ileal contents is unfavorable to the growth of the proteolytic bacteria So long as the supply of carbohydrate material is adequate the microorganisms of the acid-producing type flourish and any of the bacteria of the large intestine which may have invaded the small intestine are unable to gain a foothold Under certain circumstances, however, especially in infants, this does occur In young children, in whom the protective power of the small intestine is even less than that of adults, a severe

type of toxemia results, characterized by vomiting, diarrhea, dehydration, acidosis, fever, emaciation and great prostration

Some believe that guanidine is the toxic product concerned in certain types of intestinal intoxication in infants In intestinal disorders associated with the extension into the small intestine of the flora of the large, measures are directed toward encouraging the normal acid-producing type of organism in gaining the upper hand Carbohydrate is supplied, usually in the form of lactose, which passes farther down the intestinal tract than other sugars before it is absorbed It has also been common practice to administer cultures of the fermentative type of organism, such as that of sour milk—*Bacillus bulgaricus* (so-called as the result of the writings of Metchnikoff who attributed the health and longevity of the Bulgarian peasant to his drinking large quantities of sour milk) Today cultures of *Bacillus acidophilus*, one of the normal inhabitants of the small intestine, are usually employed Ordinary *antiseptics* given with the view of inhibiting bacterial growth exert no appreciable effect, although certain sulphonamides are valuable for their bacteriostatic action

ACUTE INTESTINAL OBSTRUCTION

Symptoms of intense severity result when the lumen of the small bowel is completely obstructed as a result of constriction by an adhesive band, kinking, twisting or pressure by new growth, intussusception, strangulation by a hernial ring, etc The condition is ushered in by severe cramp-like abdominal pain, vomiting and shock If the condition is not relieved by operation, reverse peristalsis arises above the point of obstruction, intestinal contents pass into the stomach and the vomiting becomes fecal in character Later the bowel above the obstruction loses its tone, becoming dilated and filled with intestinal secretion, and gas The loss of fluid in the vomitus and the drainage of large quantities of fluid into the distended bowel leads to a *fall in blood chloride, alkalosis and dehydration* (p 24) Other blood changes are a *rise in the non-protein nitrogen* and an *increase in the fibrinogen content* The former is the result of tissue destruction combined with impairment of renal function Great prostration occurs, ending in death

Functional obstruction The dilated, immobile state of the bowel described above as occurring in the later stages of a mechanical obstruction

occurs from other causes and is then referred to as *paralytic* or *adynamic ileus*. The most common cause of paralytic ileus is peritonitis, it may also result from some severe intestinal injury or undue handling and exposure of the bowel during abdominal operations. Reflex inhibition of the bowel wall occurs followed by distention. The distended bowel is incapable of propelling the feces along its lumen, that is, obstruction of a functional character exists. When the bowel above a mechanical obstruction has become atonic and dilated, functional obstruction tends to persist even after the mechanical block has been relieved by operation.

EXPERIMENTAL OBSTRUCTION When the intestine of an animal is tied across the symptoms which follow are chiefly weakness, prostration and vomiting. The animal shows little or no evidence that it is suffering pain. The higher in the intestinal tract that the obstruction is made, the more severe are the symptoms, and the shorter is the duration of life after the operation. Following obstruction of the colon the animal may survive for some weeks, whereas after obstruction of the jejunum or duodenum it dies as a rule within five or six days. Reduction in blood volume (anhydremia), fall in blood chloride (see below), increased alkali reserve and non-protein nitrogen of the blood, and a rise in the percentage of fibrinogen occur.

Paralytic ileus may be produced in dogs by the injection of a solution of iodine into the peritoneal cavity. It was shown by Campbell and Markowitz that the intestinal inhibition produced in this way could be abolished by spinal anesthesia which apparently blocked inhibitory impulses reaching the bowel through the splanchnic nerves (p. 584-5).

A CONSIDERATION OF THE CAUSE OF DEATH IN MECHANICAL OBSTRUCTION

A number of theories have been advanced in recent years in attempts to give an explanation for the symptoms and death in acute mechanical obstruction of the bowel. We may, at the outset, dismiss the possibility that death is simply the result of blockage of the alimentary tract, and the prevention of the passage of food along its lumen. It was shown originally by Stone, Bernheim and Whipple that if a few inches of the bowel were excised, both ends of the segment closed and the continuity of the digestive tract then re-established by an anastomosis of the cut ends of the bowel, death occurred even more rapidly than if the bowel had been obstructed by ligation. The sur-

vival time after the closure of such an isolated segment or loop is rarely more than 3 or 4 days and may be only 24 hours.

Toxic theories

A *bacterial theory* in one or other of its modifications has had its adherents. The toxin of the Welch bacillus has been suspected by some, but little support has been given by recent work to the belief that this or any other bacterial toxin is responsible for the symptoms. Others have thought that a toxic agent derived from the bacterial decomposition of protein within the intestinal lumen was the lethal agent. It is impossible to consider seriously any bacterial theory of intestinal obstruction, for such theories ignore the fact that the symptoms diminish in intensity, and the survival time is lengthened, the lower in the intestinal tract that the obstruction is produced. Bacterial growth, protein decomposition and histamine concentration on the other hand, are all greater in the lower than in the higher reaches of the intestinal tract.

The dechlorination and dehydration theory

It was first shown by Hartwell and Hoguet (1912) that obstructed animals which survived for some days showed marked dehydration. Vomiting, it was claimed, was responsible for the reduction in body water, the symptoms of obstruction and death were believed to be the direct consequence of the dehydrated state. They showed, and their observation was confirmed by Haden and Orr and others, that the life of an animal with an obstructed intestine could be prolonged by the subcutaneous or intravenous administration of saline.

That the loss of chloride in the vomitus may cause a profound fall in blood chloride, and dehydration is an established fact. Hastings, Murray and Murray found that in dogs, obstruction at the pylorus caused a reduction in the blood chloride to 50 per cent of the normal within a few days after the operation. Gamble and Ross confirmed this observation and made a more extended study of the changes in blood chemistry which follow pyloric obstruction. They consider that loss of chloride, by leading to a reduction in the electrolytes of the body, is the primary cause of the dehydration. They state that "a withdrawal of the electrolytes of the body fluids will be accompanied by a proportionate reduction in the volume of

body water and that this change can only be repaired by replacing both the lost water and the lost electrolytes." These observers showed that the loss of chloride in the vomitus is repaired for a time by the retention of carbon dioxide and the consequent increase in bicarbonate. By this means the concentration of electrolytes is maintained. In other words, the sum of (Cl^-) and (HCO_3^-) remains constant for a time (30 hours or so) after the obstruction has been established. A degree of alkalosis, however, results, which is countered by a loss of base (Na) in the urine, a reduction in $(\text{Cl}^-) + (\text{HCO}_3^-)$ occurs. The reduction in ionic concentration of body fluids resulting from the depletion of base, which unlike Cl^- cannot be substituted for, is accompanied by a proportionate loss of water (fig 43.6). The value of sodium chloride injections in prolonging life in obstructed animals depends chiefly upon the fact that the replenishment of the stores of sodium permits the retention of water. Chlorides such as NH_4Cl or KCl are without any beneficial effect, nor will water alone or a solution of glucose prolong life.

The reduction in the blood chloride is not necessarily the result of vomiting. The accumulation of secretions above the point of obstruction will just as surely cause chloride depletion and dehydration. In the rabbit for example, which cannot vomit, obstruction causes the characteristic blood changes. A most powerful stimulant to secretion is distension of the bowel wall, the dilated bowel in the later stages of obstruction thus becomes a receptacle for large quantities of fluid. In the case of the rabbit, fluid amounting to thirteen per cent or so of the body weight may be found in the stomach and bowel after death. The fatal effect of distension has been shown by Herrin and Meek. They distended a loop of bowel by means of a balloon and allowed the continuous secretion which resulted to drain to the exterior. The animals died in from 6 to 14 days. A loop of bowel opening to the exterior but not distended causes no injurious effect. Herrin and Meek concluded that the loss of chloride in the secretions was the essential factor leading to the death of their animals. Dragstedt and Ellis have also shown that the profound fall in blood chloride which follows the drainage of the gastric juice to the exterior is accompanied by grave symptoms and ultimately ends in death. The symptoms are rapidly relieved by the administration of saline. Fine and his associates have reported a pronounced reduction in the plasma

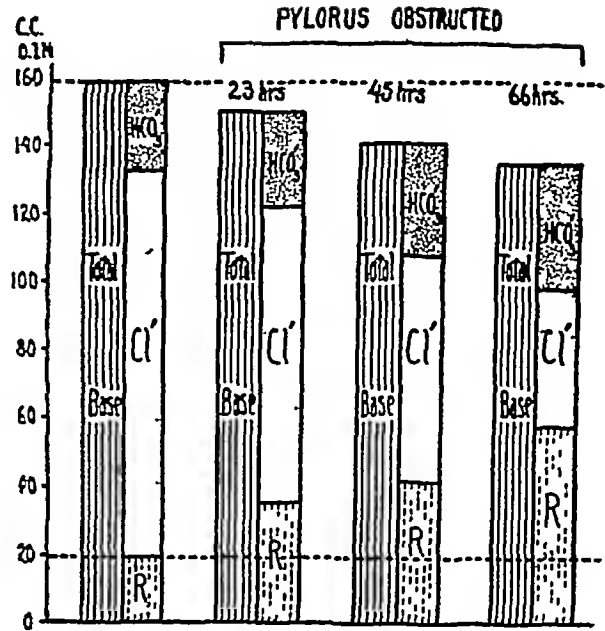


FIG 43.6 Diagram to show change in the total content and distribution of electrolytes in the body fluids following pyloric obstruction (After Gamble and Ross)

volume of patients suffering from acute obstruction of the small intestine.

There is no doubt that chloride loss and dehydration are important factors in the later stages of obstruction. The administration of saline is one of the most valuable measures possessed by the clinician for combating the condition at this stage either before or following operation.

Nevertheless, though the blood changes at this time undoubtedly prejudice the animal's chance of survival, it is unlikely that they play the primary rôle in causing death. For one thing, correction of the blood chemistry, though it prolongs life, does not permit survival. Moreover, death frequently occurs before any significant change in the blood chlorides takes place, while on the other hand, the blood chloride may be maintained at a level considerably lower than that usually seen in obstruction without the animal becoming ill. The following experiments⁷ indicate a nervous element as the primary cause of the train of symptoms seen in acute obstruction. A balloon was placed in the lower duodenum and distended to a pressure of about 100 mm Hg. A large sized rubber tube attached to one side of the balloon and extending into the bowel below prevented any actual obstruction. X-ray examination showed that a barium mixture moved freely beyond the dis-

⁷ Taylor, Weld and Harrison



FIG. 437 X-ray photograph showing Miller Abbott tube in the small intestine.

distended region. These animals had symptoms indistinguishable from acute obstruction yet little fluid was lost by vomiting and the blood chlorides showed practically no change. On the other hand, otherwise normal animals who had had the blood chloride depleted by means of histamine injections (to stimulate gastric secretion) and apomorphine (to induce vomiting) remained in apparent good health even though the blood chloride had been maintained at a level 50 per cent below the normal for three weeks.

It is evident that in the balloon experiments just mentioned, distension of the bowel caused death in some way other than by chloride depletion. Another group of animals in which the bowel was distended by a balloon as described, but in which the segment of bowel had been first denervated survived for a much longer time than those in which a preliminary denervation had not been performed. This result indicates that afferent impulses arising from the distended segment of bowel are concerned in some way with the severe symptoms of obstruction. Herrin and Meek also observed that if a distended draining loop were

denervated the animals survived indefinitely. They, however, offer an explanation of the benefit derived from denervation based upon the theory that dechlorination plays the essential rôle, namely, that when the vomiting and loss of appetite produced reflexly from the distended loop are abolished by nerve section, chloride loss is reduced and chloride ingestion increased. The importance of afferent impulses from the intestine in the production of the symptoms of obstruction therefore must be admitted, that they may cause death quite independently of dechlorination is undoubted. It is well known that many reflex effects upon the cardio-vascular system and upon several important functions can be initiated from the gastro-intestinal tract. The severe pain and collapse which result from the distension of other hollow viscera, e.g., the gall-bladder, stomach, kidney pelvis, etc., may also be recalled in this connection. Moreover, the severe symptoms and early death which result from the mere distension of the bowel by a balloon calls to mind the analogous clinical condition known as Richter's hernia. In this condition only a portion of the intestinal lumen is pinched off and isolated from the main passage. The bowel is not obstructed, yet all the symptoms of acute obstruction are present. It is quite conceivable that acute distension of the small isolated pocket by secretions and gas is responsible for the symptoms.

The importance of distension as a factor in acute obstruction of the intestine is now generally recognized and it has become the practice in suitable cases to decompress the bowel by suction-drainage through a tube passed from the mouth to the duodenum (Wagensteen and Paine) or, as in the method of Miller and Abbott, to insert a long slender tube through the nose into the stomach and allow it to be carried to the level of the obstruction. The tube has a double lumen and is provided at its tip with a small balloon which can be moderately inflated after it has entered the duodenum. The balloon serves as a stimulus to peristalsis which carries it and the tubing along the bowel to the point of obstruction (fig. 437).

CHAPTER 44

VISCERAL SENSATIONS

The abdominal and thoracic viscera are insensitive to the several types of stimuli which readily arouse sensations in the skin and more superficial tissues of the body. The effects of the different varieties of stimulus—*thermal, chemical, tactile* and *pain*—have been investigated by a large number of observers.

Sensations of pain cannot be elicited from the viscera by the *usual means*. The intestine or the liver, the stomach or the heart may be cut, burned or pinched without arousing any immediate sensation. In the second stage of a colostomy operation, for example, the colon can be opened without pain being experienced by the patient. Harvey remarked upon the absence of sensation in the exposed heart of Viscount Montgomery.¹

The insensibility of the alimentary tract to ordinary forms of stimulation commences in the lower or middle third of the esophagus and extends as far as the commencement of the anal canal. How can these observations be reconciled with the well-known fact that pain is one of the commonest manifestations of visceral disease?

The whole subject of pain arising in or referred from the viscera is still highly controversial. No final answer can be given. The most that can be attempted is a summary of some of the more significant experimental results and clinical observations, and the opinions of those who have especially interested themselves in the subject.

Lennander considered that abdominal pain was always due to the stimulation of somatic nerve terminals in the parietal peritoneum or root of the mesentery; the bowel itself was supposed to be quite devoid of pain fibers. According to this view therefore, pain localized within the organ itself—*true visceral pain*—was an impossibility.

Ross's theory postulated that pain from the abdominal viscera was of two types: (a) *referred*

(*somatic*) *pain*, and (b) *true visceral* or *splanchnic pain*, which was diffuse, poorly localized but felt in the viscus itself.

Mackenzie, as a result of his clinical observations, strongly supported Ross's idea of referred pain, but maintained that *all* visceral pain was of this nature, the viscera being quite insensitive to all forms of stimulation. In agreement with Lennander, he did not believe that the viscera contained pain fibers; true visceral pain, i.e., pain in the organ itself, was therefore never experienced.

Conception of referred pain according to Ross and Mackenzie

When a viscus is diseased, pain or tenderness is frequently felt in the tissues overlying it approximately (abdominal or chest wall) or in some part quite remote from it. For example, pain is felt in the neck or shoulder (fig. 44-1) in conditions affecting the diaphragm, between the scapulae in gastric disease, in the region of the umbilicus in appendicitis, in the testis in renal colic and in the sternal region, or down the left arm in angina pectoris (p. 328). When traction is made upon a coronary artery of a dog, the animal whines and indicates the location of the pain by limping on the left fore-paw. In dextrocardia anginal pain is felt in the right arm. Other examples are, the pain in the perineum and tip of the penis caused by a stone in the region of the neck of the bladder, and the pain in the groin due to a stone in the ureter.

It will be recalled (ch. 66) that a given spinal segment supplies a visceral area with autonomic nerve fibers (afferent and efferent) and also a well-delineated area of the skin (dermatome) with somatic nerves. The two types of structure linked in this way through the afferent nerves and the central nervous system may be some distance apart (e.g., diaphragm and shoulder) or be more closely related (e.g., area of abdominal wall and an underlying abdominal viscus). Mackenzie believed that afferent autonomic impulses arising in a diseased organ, though of themselves incapable of arousing any sensation, would upon entering the spinal cord set up an "irritable focus" with the result that cells accustomed to receive impulses from the corresponding somatic area were excited. Thus, the

¹ Harvey records, "I carried the young man to the King (Charles I) that His Majesty might with his own eyes behold this wonderful case, that, in a man alive and well, he might, without detriment to the individual, observe the movement of the heart, and with his proper hand even touch the ventricles as they contracted. And His Most Excellent Majesty, as well as myself, acknowledged that the heart was without the sense of touch, for the youth never knew when we touched his heart."

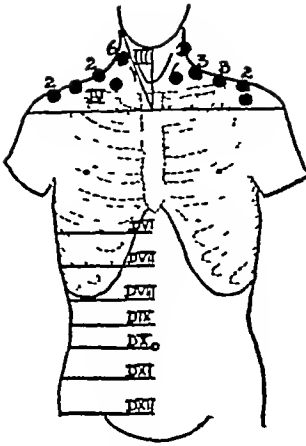


FIG 44.1 Location of maximum points of referred neck pain from irritation of the diaphragm. The figures refer to the number of stimulations in each instance. They are all situated in the region supplied by the third and fourth spinal segments. (After Capps)

impulses from the viscera "irradiated" are spread on to cells of the corresponding somatic center. New impulses originating in these cells travelled along the usual paths to higher perceptive centers (thalamus) which projected or referred the sensa-

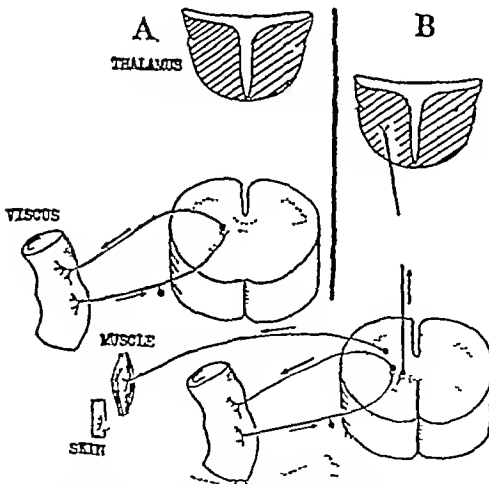


FIG 44.2 Diagram to illustrate Mackenzie's theory of referred pain. A, representing normal conditions, a visceral reflex is shown. B, illustrates a visceromotor reflex. Impulses from a viscus are conceived as setting up an irritable focus in the cord which through the involvement of neighboring neurons increases the tone of muscles innervated by the same segment, and causes a discharge of impulses over the pathway for pain. The sensation is then projected in consciousness to the periphery, as indicated by the dotted line.

tion to the somatic area e.g., skin or muscle, from which it was accustomed to receive impulses. In this way, *spontaneous pain* in superficial structures remote from the diseased site was accounted for. Mackenzie spoke of these reactions as *viscero-sensory "reflexes"*.² *Tenderness* to touch, pressure or light pinching of the skin (hyperesthesia and hyperalgesia) was ascribed to the impulses, which ordinarily would be below the threshold for pain, arriving in the segment rendered hyperexcitable as a result of impulses received from the diseased organ. In the case of the abdominal viscera, Mackenzie claimed that the area of tenderness in the abdominal wall remained fixed, though the position of the diseased organ changed, thus indicating the referred nature of the pain. He explained the rigidity (hypertonus) of muscles overlying a diseased organ, the right rectus abdominis in acute appendicitis, for example, upon a similar basis. Afferent impulses of normal intensity arising in the muscle proprioceptors upon arriving at the spinal centers, which had been rendered hyperexcitable by the receipt of abnormal visceral impulses, resulted in a reflex increase in tonus of the corresponding muscles. The referred motor reaction he spoke of as a *viscero-motor reflex* (see fig 44.2).

The conception of referred pain was supported by the work of Head, who mapped out the segmental distribution of the cutaneous nerves responsible for hyperalgesia in diseased states, and showed that they came from the same segments as received autonomic fibers from the diseased organ.

In table 42 are given the segmental areas to which pain is referred in disease of various viscera (Head).

True visceral (splanchnic) pain

True visceral pain, though denied by Mackenzie, exists. The pains of dysmenorrhea or childbirth, for example, or the pain of intestinal colic, of an over-distended bladder, or the substernal pain of coronary occlusion, though diffuse, seems to arise in the organ itself, even when referred as well, to a somatic structure, such as muscle or skin. Kinsella observed that, in conscious patients, pressure with the fingers upon an inflamed appendix caused pain centered in the viscus; a healthy appendix is insensitive. He and others have also shown that a chronic duodenal ulcer is also tender to direct

²This is obviously a most unsuitable term for it is not a reflex in the ordinary accepted sense of that word.

TABLE 42*

VISCUS	SPINAL SEGMENTS
Lungs	1-7 dorsal, mostly 2-5 dorsal
Heart	3-5 cervical, 1-8 dorsal, predominantly on left side, sometimes bilateral
Esophagus	Mainly 5 dorsal, also 6, 7 and 8 dorsal
Breast	4 and 5 dorsal
Stomach	7, 8 and 9 dorsal, usually bilateral
Intestine	9-12 dorsal, bilateral or on left side only
Liver	8-10 dorsal on right side
Gall bladder	Mostly 8 and 9 dorsal, also 5-7
Kidney	Mostly 10 dorsal, also 11 and 12 dorsal and 1 lumbar
Ureter	11 and 12 dorsal and 1 lumbar
Testis	10 dorsal
Epididymis	11 and 12 dorsal
Bladder	11 and 12 dorsal and 1 lumbar, also 3 and 4 sacral
Prostate	10 and 11 dorsal, also 1-3 and 5 sacral
Ovary	10 dorsal
Fallopian tubes	11 and 12 dorsal
Uterine cervix	11 and 12 dorsal and 1-4 sacral
Uterine body	10 dorsal to 1 lumbar

*From W R Brain, Diseases of the Nervous System, after Head

contact In transection of the cord at the third thoracic vertebra in a patient reported upon by Kinsella in which there was complete paralysis and anesthesia of the abdominal wall, intestinal pain was experienced, which obviously could not be caused by reference to somatic nerves Impulses in such cases must reach the central nervous system along intact visceral (autonomic) afferents They travel apparently by sympathetic fibers which enter the cord above the level of the transection It is unlikely that they are transmitted by the vagus

Pain in a normal viscus may also result from disease in a distant organ through a visceromotor reflex Painful pylorospasm, for example, may accompany disease of the appendix A distinction should be drawn between pain of this nature and referred pain

The adequate stimulus for true visceral pain

The existence of true visceral pain is not incompatible with the statement made above that the viscera are insensitive to the ordinary types

of stimulus From his investigations Hurst concluded that the only adequate stimulus for visceral pain fibers is tension Distention of a hollow viscus, e g, stomach, intestine, gall-bladder, etc, gives rise to pain as a result of the stretch stimulus applied to the nerve terminals in its wall The pain is roughly localized to the viscus itself, or referred Under certain circumstances chemical substances produced in an ischemic organ, as in angina pectoris or intermittent claudication, may stimulate nerve fibers subserving pain

Though the question is by no means settled, it appears that *contraction*, alone, of the muscular wall of a hollow viscus, such as the intestine, does not cause pain unless the movement of the bowel makes traction upon the mesentery Pain arises, however, if the contraction causes distension of a neighboring portion of the wall as may result when the contraction wave approaches a mechanical obstruction, a length of bowel in spasm or a sphincter which fails to relax (*achalasia*) Poulton, for example, found that when a balloon was inserted into the lower part of the human esophagus, the approach of a peristaltic wave toward the obstruction caused pain, but during its passage over the esophageal wall in contact with the balloon no sensation was felt Poulton ascribes the absence of pain during the passage of the contraction wave to the reduction in the diameter of the tube and the consequent release from stretch of the nerve endings lying between the muscle fibers Pain also ceased if the esophageal muscle relaxed to accommodate the balloon, that is, adjusted the length of its fibers to the distending force (postural tone, see p 557) As further evidence for the effectiveness of distension in causing pain, the following observations may be cited In animals, when an intestinal loop exposed under local anesthesia is stimulated to powerful contraction, there is no evidence of pain, whereas even moderate distension of the loop (as by inflating it with a balloon) is manifestly painful Distension of the gall-bladder of the cat is accompanied by reactions indicative of intense pain Distension of the human appendix by the injection of fluids through an appendicostomy opening causes severe pain in the epigastrium or in the region of the umbilicus, and when the duodenum is distended by the injection of material through a duodenal tube, pain is felt on the right side The pain impulses undoubtedly reach the central nervous system through the splanchnic nerves Bentley and

Smithwick distended the duodenum of patients whose splanchnics had been divided on one or on both sides. After unilateral denervation, distension caused pain on the denervated side only, no pain whatever was felt after bilateral splanchnic section.

Morley's theory of the parietal origin of referred abdominal pain

Morley contends that the referred pain of abdominal disease does *not involve autonomic afferents* as postulated by Mackenzie but is due to the stimulation of *somatic* pain fibers in the parietal peritoneum or mesentery, the sensation being referred to the superficial area innervated from the same spinal segment. He expresses what he terms the *law of referred pain* in the following words: *Referred pain only arises from irritation of nerves which are sensitive to those stimuli that produce pain when applied to the surface of the body*, that is, by the stimulation of somatic sensory nerves. He believes that the somatic innervation of the peritoneum extends along the mesentery to within a short distance of its attachment to the bowel and does not terminate as has been generally supposed near the root of the mesentery.³ The mesentery or peritoneum is therefore sensitive to tearing, cutting, etc., whereas the organ itself which contains only autonomic afferents is insensitive to these types of stimulus. In support of his views Morley cites the pain in the shoulder associated with irritation of the peritoneal covering of the diaphragm. He points out that the diaphragm is innervated (through the phrenic) chiefly by the fourth cervical spinal segment and to a less extent by the third and fifth. None of these segments give rise to autonomic fibers but the third and fourth cervical nerves furnish somatic afferent fibers to the shoulder area in which the referred pain of diaphragmatic disease is located. Morley explains muscular rigidity in a similar manner—the radiation of impulses over the motor nerves.

³ Sheehan has made a study of the nerves of the mesentery and finds the following types of fiber:

(1) Fibers ending in Pacinian corpuscles scattered throughout the mesentery. These are afferent sympathetic fibers which travel in the splanchnic nerves.

(2) Free non myelinated fibers. These are afferent and efferent sympathetic fibers, their terminals are distributed to the serous covering of the bowel. They provide a medium for the transmission of true visceral pain.

(3) Free myelinated fibers derived presumably from somatic nerves. These apparently do not extend as far as the serous covering of the bowel itself.

He replaces Mackenzie's terms, viscerosensory and visceromotor reflexes by *peritoneo cutaneous radiation* and *peritoneo muscular reflex* respectively. Morley also recognizes spontaneous, true (unreferred) visceral pain, resulting from an adequate stimulus—namely, tension. The pain and tenderness resulting from pressure upon an inflamed viscus (e.g., ulcer of the duodenum) through the abdominal wall, and which seems to arise in the organ itself, he ascribes however, to the parietal peritoneum being brought into contact with the roughened surface of the lesion. As evidence that the sensation is not, in such instances, referred to the skin from the diseased structure, but is due to the stimulation of nerves in the parietal peritoneum he states that (a) the area of tenderness shifted with the movement of the viscus, and (b) direct pressure upon the inflamed organ when exposed by operation in a conscious patient (i.e., under local anesthesia) did not give rise to any sensation (see fig. 44-3).

Morley has made out a case for the production of pain in some instances through a peritoneo-cutaneous reaction, and for pain and tenderness upon pressure being the result of the stimulation of somatic nerves in the parietal peritoneum. But his views cannot be accepted in their entirety. Others have reported that a duodenal ulcer is tender to direct touch (p. 598), and there is no doubt that referred pain can be brought about through the mediation of either visceral or somatic nerves. Woollard and Carmichael, for example, have obtained evidence for the latter from experiments upon the human testis which, since it has migrated from the abdominal cavity and is enveloped by peritoneum, may be looked upon as an abdominal organ, in so far as the question of referred pain is concerned. These observers found that after all the nerves to the testis had been blocked by means of novocaine, except the autonomic fibers passing along the spermatic artery, no sensation was felt within the organ when it was compressed, but pain, referred to the tenth dorsal segment—lower abdomen and back—was experienced. The observation of Bentley and Smithwick, mentioned above, point in the same direction.

The views of Lewis and Kelgren

Lewis and Kelgren postulate a common system of afferent nerves supplying deep somatic structures as well as the viscera. They maintain with

good reason that since referred pain can follow the irritation of either autonomic or somatic afferent nerves, there is no physiological justification for making a distinction between pain referred from a viscus and that referred from the parietes, nor between the motor reactions (e.g., muscular rigidity) resulting from stimulation in either situation. Though in the first instance autonomic afferent fibers are stimulated, and in the second it is somatic sensory nerves, the impulses in either case enter the spinal cord by the posterior nerve roots, and the nerve fibers which transmit them have their cell-stations in the posterior root ganglia, there is no indication that they do not follow identical paths in the central nervous system. Thus impulses from a viscus or a deep somatic structure cause a sensation which in either case is projected to a part remote from that stimulated but innervated by the same spinal segment. The pains whether referred from a viscus or from part of the soma, have common characteristics, being diffuse and poorly localized.

Lewis and Kellgren carried out experiments on a number of human subjects, their results are highly significant to the question of referred pain. Pain closely resembling in character the referred pain of visceral disease was induced by the injection of a small quantity (0.3 cc) of hypertonic saline into an interspinous ligament at various spinous levels. Injection into the 1st lumbar interspinous ligament caused pain distributed in a manner strikingly similar to the pain of renal colic, namely, in the loin and in the inguinal and scrotal regions. The pain was accompanied by retraction of the testis. Injection into the 9th thoracic interspinous ligament caused pain in the back in the region of the 1st lumbar spine and over an area in front extending from the 9th costal cartilage on the affected side to the umbilicus. Rigidity of the abdominal muscles and deep tenderness were associated with the pain. Stimulation of the 8th cervical ligament was followed by pain in the interscapular region, over the pectoralis major muscle and down the inner side of the elbow and forearm, together with a sensation of constriction in the upper part of the chest on the stimulated side. Several subjects of angina pectoris were chosen for experiment and were asked to compare the pain which they experienced in an attack with that caused by the stimulation of the 7th cervical or the 1st thoracic interspinous ligament. In all instances the patients described the experimentally induced pain as being identical in character with that caused by the disease, though it showed some minor differences in distribution.

Lewis and Kellgren also found that in cats mechanical stimulation (pinching) of the pancreas or of the mesentery in the duodenal loop caused motor reflexes

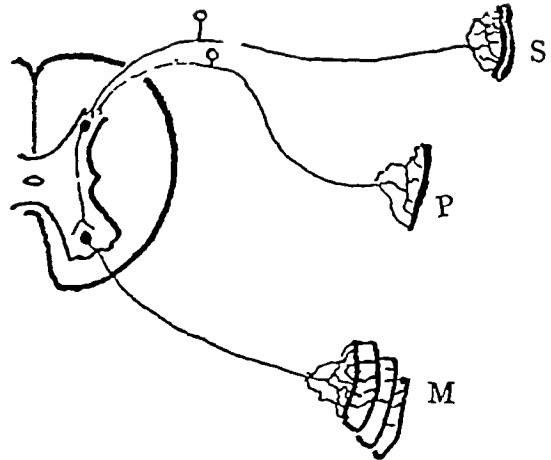


FIG 44.3 Diagram to illustrate Morley's theory of peritoneocutaneous radiation and peritoneomuscular reflex, M, muscle, P, peritoneum, S, skin. After Morley, redrawn.

from the abdominal muscles which resembled closely those caused by stimulation of the back muscles.

An explanation of referred pain on the basis of reflex vasomotor changes (viscero-somatic reflex) or the liberation of a chemical substance at somatic nerve endings

Pollock and Davis stimulated the peritoneal surface of the diaphragm in dogs. The animals showed every sign of suffering pain, which was abolished by any one of the following procedures: Section of the phrenic nerve, removal of the cervical sympathetic chains, severing the eighth cervical and the first, second and third thoracic anterior roots, transection of the cord at the seventh cervical segment, destroying the cord at the first and second dorsal segments, or section of the cervical posterior roots. Pollock and Davis conclude from these results that the pain impulses following stimulation of the diaphragm travel over the phrenic, enter the cord by the posterior cervical roots, descend the cord to the level of the eighth cervical and first, second and third thoracic segments. Connections are then made with cells in the lateral horn of gray matter (intermediolateral column) from where impulses pass by sympathetic preganglionic fibers to the cervical sympathetic chain, and then by postganglionic fibers "effectors in the skin, blood vessels, meninges and other structures, where by some vasomotor (?) or hormonal (?) process the sensory endings of the cerebro-spinal system are stimulated and a sensory impulse travels over the ordinary cerebro-spinal system, enters the spinal cord by the posterior roots and ascends to consciousness." The pain and cutaneous lesions of herpes zoster come to mind in this connection.

The diffuse character of deep pain in contrast to the accurate localization of cutaneous stimulation can be accounted for by (a) the existence of

relatively few afferent fibers in deep structures, (b) the lack of experience and training in the localization of sensations from the viscera, since such are normally felt very infrequently, and (c) when impressions from the internal organs do occur their localization is not aided by sight as in the case of more superficial sensations. However, in experiments with intelligent subjects having some knowledge of anatomy, and in whom pain was evoked by inflating balloons inserted into hollow viscera, the subjects' sense of localization was much improved above the ordinary

THE SENSIBILITY OF THE ALIMENTARY TRACT TO TACTILE, THERMAL AND CHEMICAL STIMULATION

Touch The sensation of touch disappears at the lower end of the pharynx. This was shown by Hurst by means of an esophageal tube with a slit on one side through which the mucosa of the esophagus could be stimulated. Carlson, by means of a test-tube brush passed into the stomach, was unable to elicit the sensation of touch from the gastric mucosa. The rectum possesses no tactile sensibility but the anal canal is sensitive.

Thermal sensibility The esophagus is sensitive to extremes of heat and cold. The sensations of temperature that are felt when hot or cold materials enter the stomach have been thought by some to originate in the lower end of the esophagus, by others to arise in the skin of the epigastrium, either through thermal conduction or by reflex changes in the cutaneous blood vessels. Carlson has shown, however, that the gastric mucosa is sensitive to extreme temperature changes, i.e., protopathic thermal sensibility (below 13°C or above 45°C). It is the lower end of the esophagus, however, which is responsible for the greater part of the thermal sensation that is experienced when excessively hot or cold materials are swallowed. This is due to the greater sensitivity of the esophageal mucosa, as well as to the fact that the material is retained for an appreciable length of time above the cardia. The colon is insensitive to temperature changes, but even comparatively slight differences in temperature can be detected in the anal canal.

Chemicals, with the exception of alcohol, cause no sensation whatever when introduced into the stomach or intestinal canal. The mucosa is completely insensitive to acids, the introduction into the healthy stomach of a solution of 0.5 per cent hydrochloric acid causes no pain or sensation of any kind. Alcohol stimulates the mucosa of the esophagus and stomach and causes a sensation of warmth. Peppermint and various condiments free from alcohol arouse no sensation. The pelvic colon and rectum show a similar sensitivity to alcohol but are insensitive to other chemicals. The anal canal is extremely sensitive especially to alcohol and glycerine. Both cause a burning sensation.

NAUSEA

Nausea usually precedes the act of vomiting (p 572) but may occur alone. On the other hand vomiting may occur without nausea, as in certain cerebral conditions. The sensation is felt in the back of the throat or pit of the stomach, and in its milder degrees is merely a "sinking" sensation in the epigastrium. It is frequently associated with vasomotor disturbances and sweating. Increased tension upon the walls of the stomach or duodenum is a potent cause of the sensation, and Poulton has shown that it is also induced by distension of the lower part of the esophagus. During the passage of a peristaltic wave which relieves the tension upon the nerve fibers (p 599) in the esophageal walls the sensation is relieved. Barclay showed by radioscopy in the human subject that nauseous odors caused the lower border of the stomach to descend an inch or two evidently as a result of sudden relaxation of the abdominal muscles. This movement would tend to stretch the esophagus and gastric walls and so exert tension upon the nerve endings. The stimulus which induces nausea is, therefore, the same, apparently, as that which causes visceral pain, but of lower intensity. It is likely that the sensations experienced during changes in speed of an elevator are also the result of tension exerted upon the esophagus and gastric walls. This element is also probably a contributory factor in the production of sea sickness, being brought into play by the pitch and roll of the ship (p 574).

The relief of nausea and vomiting by the application of counter irritants to the epigastrium or over the sternum is probably due to either a reflex change in the postural tone of the gastric walls or to the reflex initiation of peristaltic contractions (p 599). The tension upon the nerve endings in the latter instance is taken up by the muscle fibers. Poulton observed, for example, that the sensations caused by a balloon in the esophagus were relieved by vigorous friction of the skin over the sternum, contraction of the esophagus in some cases or adjustment of the postural tone of the esophageal wall in others were observed to accompany the disappearance of the sensation.

HUNGER AND APPETITE

Hunger is a complex sensation to account for which more than one theory has been proposed, it is the principal factor controlling food intake and, consequently, body growth and nutrition. To a large extent local elements (gastric and duodenal) enter into the hunger sensation. Strong peristaltic

contractions of the gastric walls—hunger contractions—and a sense of hunger are frequently associated, an association which was clearly shown by Cannon and Washburn and by Carlson. Filling the stomach with bulky material that possesses little nutritive value, or is indeed totally inedible, inhibits the contractions characteristic of the empty organ, relieves the pangs of hunger and may even create a feeling of satisfaction. This fact has some practical bearing in the treatment of obesity, since the subject may satisfy himself with foods such as salads, fruits, etc., possessing large volume relative to their caloric values.

The contractions of the empty stomach are often accompanied by hypoglycemia. This led to the conclusion that the blood sugar level was the fundamental factor governing the sensation of hunger or of satiety. The hunger induced by insulin appeared to support this concept. But this theory cannot be sustained in the face of the increased hunger of diabetes mellitus and of starvation or hunger diabetes, in which a subject starved for some time previously continues to consume food and feels hungry after the blood sugar has risen well above the normal level. Nor has it been possible to establish any correlation between the hunger sensation and spontaneous variations in the glycemic level.

Though it cannot be denied that purely local elements—gastric or duodenal—contribute toward the hunger sensation, or the “feeling of emptiness”, there is much evidence for a general tissue or central factor. As mentioned above, hunger can be appeased by filling the stomach with material of little or no food value, but Adolph found that the diminution in food intake of rats caused by adding roughage to the diet was only transitory, the amount which the animals ate was regulated, it seemed, mainly by the potential energy value of the diet, and was dependent to a relatively small extent upon “local conditions” of the gastrointestinal tract. The introduction of nutrients by a route other than the mouth inhibited the urge to eat. The selection by animals of an essential dietary constituent, a deficiency of which had existed previously, also points to a general body state or some central mechanism as controlling the urge to eat. Finally, complete gastrectomy and denervation of the gastrointestinal tract does not abolish hunger.

In a prolonged fast in man the hunger sensation is most intense during the first few days and, though it abates after this, usually persists (con-

trary to the general belief) to the end of the fast (Carlson and Hoelzel).

The glucostatic theory of hunger, a theory in which the utilization of glucose by the tissue cells rather than the blood sugar level is the factor controlling food intake, has been proposed by Mayer. Thus the hunger and hyperglycemia of diabetes are reconciled on the basis of a low glucose utilization in this disease. Arterio-venous glucose differences were determined in human subjects and taken as a measure of glucose utilization. A-V glucose values (low glucose utilization) under 15 mg per 100 cc. were associated with a low caloric intake and hunger. A high glucose A-V difference (over 15 mg per cent) was found when the caloric intake was adequate and the subject's hunger was appeased. In diabetes mellitus and in hunger diabetes the glucose A-V values were near zero. Insulin administration caused at first a rise in the glucose A-V difference, but a compensatory fall occurred later. This theory postulates the dependence of the control of food intake and the sensation of hunger upon cells (glucoreceptors) in the central nervous system, most probably in the hypothalamus, which arouse the urge to eat when the utilization of glucose falls below a certain level. Evidence for a “feeding center” of some sort in the hypothalamus is furnished by the demonstration that a lesion placed in the medio-ventral hypothalamic nuclei causes an increase in food intake, a lesion placed more laterally causes the opposite effect, the feeding urge is depressed. A higher center appears to lie in the prefrontal area of the cerebral cortex, which is connected through thalamic nuclei with the hypothalamus (ch 67). In man, lesions of this region are not uncommonly associated with abnormal variations in the food intake.

Appetite, on account of its close association with hunger, and its vague localization, is a difficult sensation to define, some persons refer it to the pharynx or esophagus, others to the stomach. It is probably to a large degree acquired rather than inborn like the hunger sensation. That is to say, it is dependent upon previous experience, and conditioned stimuli (ch 69) enter largely into its causation. A new-born infant experiences hunger but probably not appetite. The psychic element in appetite is indicated also by its highly selective nature. A nutritious item of food, if disliked, may be rejected by a hungry person and may even drive away his appetite. On the other hand hunger whets appetite and a particular article of diet may be enjoyed which, in a state of satiety, would be

unpalatable. Nevertheless, it appears that there is also a gastric element in the production of appetite. For example, alcohol, by stimulating the gastric mucosa, causes a sensation of warmth within the stomach, raises gastric tone and arouses appetite. According to Carlson hydrochloric acid does likewise, and the flow of gastric juice induced by the taste of food may well serve the purpose of augmenting the appetite in the initial stage of digestion. It is well known that appetite is often greatly increased after the first few mouthfuls have entered the stomach, and though the effect produced by the taste of the food is difficult to exclude, the fact points to a gastric element. Pavlov has reported, as his own experience, the pronounced effect which a glass of wine, after reaching the stomach, had in restoring the lost appetite following an illness. There is also some evidence that impulses arising not only from the mucosa but from the deeper structures of the gastric walls play a part and that increased muscular tone is a factor in the production of appetite. The smell of a disgusting material may, as we have seen (p. 602) cause a sudden lowering in the position of the stomach, a drag upon its walls, and a sensation of nausea—the antithesis of appetite. Conversely an increase in the tone of the gastric muscle, by reducing the stretch upon the nerve terminations of the stomach wall, would conceivably be conducive to gastric comfort and well being, leading to the appetite sensation. In fasting the gastric tone is also high and this may well be the cause of the keen appetite usually associated with hunger. Support for the conception that gastric tone is an important factor in the production of appetite is offered by X-ray examination of the stomach in various conditions associated with disorders of appetite. In those diseases in which a poor appetite is a prominent feature the gastric tone is low, and in others characterized by excessive appetite, the stomach is very frequently hypertonic. The association of loss of appetite and gastro-intestinal hypotonicity in B_1 deficiency (p. 750) may also be recalled in this connection. Barclay has recorded experiments upon hospital patients in whom the sight or smell of appetizing food was responded to by an increase in gastric tone. In some the mere suggestion of a glass of beer was sufficient to bring this about. On the other hand, mental states, e.g., anxiety, worry, fear, etc., which lower gastric tone, also depress the appetite. The probability of appetite having its origin, in part at least, in the walls of the stomach

is indicated by what Ryle considers a very significant fact, namely, that in diffuse carcinoma of the stomach ("leather bottle" stomach) loss of appetite is a prominent and almost constant symptom. In other types of gastric carcinoma, even though far advanced, the appetite may be retained. Bitters and other so-called appetizers, unless made with alcohol exert no effect upon the appetite.

Anorexia nervosa. A mental state leading to complete loss of appetite, or what really amounts to a morbid distaste for food of all sorts, was first described by Sir William Gull some 65 years ago and named *anorexia nervosa*. The subjects are usually nervous "highly strung" women. Some emotional upset may precipitate the condition. Extreme emaciation, low metabolic rate, moderate hypoglycemia and amenorrhea are among the chief features. The appearance of the patient may suggest pituitary cachexia (Simmonds' disease, p. 801). Profound exhaustion results and death may occur from starvation. The aversion to food appears to be purely psychic, there being no evidence that some endocrine function is primarily at fault, nor is there any indication of organic gastric disease.

Bulimia, or abnormally great hunger, as already mentioned, is a symptom sometimes associated with lesions of the prefrontal part of the cerebrum.

THIRST

The sensation of thirst is referred to the pharynx and is due to the stimulation of sensory nerve endings in this situation. Two theories have been advanced to explain the mechanism by which the sensation is aroused.

According to one view, thirst is due simply to the drying of the pharyngeal mucous membrane, the salivary glands being given a role in the regulation of the water balance of the body. When the water content of the body falls below a certain level salivary secretion is depressed, the consequent drying of the mucous membrane of the throat then elicits the characteristic sensation. If such a view is correct, drying of the pharyngeal mucosa from whatever cause should cause thirst. Cannon, who provided the principal support for this theory, found in studies upon himself that after abstinence from fluids for a time the depression of salivary secretion which resulted was definitely associated with thirst. Atropine, which inhibits salivary secretion, also produced the typical sensation, and thirst aroused by the deprivation of water was relieved by the application of cocaine

to the mucosa. Pilocarpine or acid substances which stimulate the flow of saliva also relieve thirst. In dogs, however, atropine and pilocarpine are without effect upon the water intake.

According to the other view, thirst is a sensation resulting from changes in blood composition, probably to a rise in its osmotic pressure, which stimulates the afferent nerve endings, or acts perhaps upon central nervous structures. Rowntree and his associates, for example, found that the thirst of diabetes insipidus was relieved neither by cocainization of the mucous membrane nor when salivation was induced by pilocarpine. The experiments of Gilman suggest that cellular dehydration rather than a rise in osmotic pressure is the prime factor in arousing thirst. Elevation of the osmotic pressure of the blood of dogs by the injection of a hypertonic salt solution, and the imbalance thus caused in the osmotic relationship between intra- and extra-cellular fluids, caused a much greater intake of water than an equivalent rise in osmotic pressure resulting from the administration of urea. After the injection of salt, sufficient water was drunk to quickly reduce the osmotic pressure of the blood, whereas after urea, to which the cells are readily permeable, and which, therefore, did not disturb the osmotic relationships, the water intake was but little increased and the osmotic pressure of the blood remained elevated. In support of the view that a lowered water content of the cells is the true thirst stimulus, Gilman cites an experiment in which anhydremia was induced by the withdrawal of large quantities of extra-cellular electrolytes (e.g., NaCl) without the with-

drawal of water. In such anhydremic animals dehydration of the tissue cells does not occur, and though the oral mucous membranes are quite dry there is no evidence of thirst—water is refused.

Neither theory alone explains all the experimental facts. It appears that the factors concerned in the production of thirst are both local and general. The importance of a rise in the osmotic pressure of the blood in causing the sensation is suggested by the experiments of Verney on water diuresis (p. 455), for it is not unreasonable to suppose that thirst and the control of the water balance of the body through the anti-diuretic principle of the pituitary are parts of the same mechanism. As mentioned above, the intravenous administration of urea is much less effective in inducing thirst than is NaCl, and Verney found that this substance was ineffective as an inhibitor of water diuresis.

Experiments upon dogs in which the salivary glands were extirpated might be expected to provide a decisive answer to the question. Such procedures however, have yielded conflicting results. Though some observers have reported an increased water consumption after the operation (Gregersen and Cannon), others have been unable to obtain any evidence that removal of the glands induced thirst (Montgomery)⁴. A positive result, nevertheless, carries more weight than a negative one since mucus-secreting glands are scattered diffusely over the mucosae of the buccal and pharyngeal cavities.

⁴ See Montgomery, M. F., and Gregersen, M. I. and Cannon, W. B.

SECTION VI. METABOLISM AND NUTRITION

By N B T

(Exclusive of Chapters 49 and 50)

CHAPTER 45

GENERAL METABOLISM

Metabolism is the term employed to embrace the various chemical processes occurring within the tissues upon which the growth and heat production of the body depend and from which the energy for muscular activity and for the maintenance of vital activity and for the maintenance of vital functions is derived. *Catabolism* is the term applied to those reactions which involve the breakdown or decomposition of substances into their simpler constituents. *Anabolism* is the word used to connote building up or assimilative processes.

HISTORICAL SURVEY AND GENERAL PRINCIPLES

The modern science of metabolism may be taken to date from the experiments of Lavoisier, carried out toward the end of the eighteenth century. He demonstrated that animal heat was the result of the oxidation of carbon in the body, and compared the process to the burning of a candle or any other combustible material. The process in either case involved the consumption of oxygen and the formation of carbon dioxide. Lavoisier and the physicist Laplace placed a guinea pig in a closed chamber and determined the quantity of carbon dioxide eliminated in a 10-hour period. The same quantity of carbon dioxide was found to be produced by the animal's body as when 3.3 grams of pure carbon were burned in air. The guinea pig was next placed in a closed space surrounded by ice. The heat given out by the animal's body during a 10-hour period was calculated from the quantity of ice which had melted (fig. 45.1). Three and three tenths grams of carbon burned in a similar ice calorimeter were found to generate an amount of heat comparable with that given off by the animal. It was therefore concluded that the heat generated by the animal and the carbon dioxide eliminated by the lungs were the result of the combustion of approximately 3.3 grams of the body's carbon. Thus, the parallelism between the amount of heat generated in the animal body and the quantity of carbon dioxide eliminated was demonstrated.¹ A few years earlier

Crawford, using a water calorimeter, had demonstrated a parallelism between the oxygen used and the heat generated. He found that a guinea pig gave out a definite quantity of heat for every 100 ounces of oxygen which it consumed. When this quantity of oxygen was used in the burning of carbon outside the body, approximately the same quantity of heat was produced.

A summary of the work of later investigators and of the fundamental principles upon which modern studies in metabolism are based will now be given.

(1) Heat values of the foodstuffs

About half a century after the experiments of Lavoisier the work of Joule (1842) demonstrated the mechanical equivalent of heat, and Mayer and Helmholtz (1845) discovered the law of the conservation of energy which states that the sum total of energy in the universe remains constant, but that one form of energy, potential, mechanical, electrical, etc., may be converted into another. Through the work of Voit, Pettenkofer and Rubner this law was shown to hold true for the animal body. Rubner determined the heat produced by the three types of food when burned outside the body in a calorimeter and thus established their heat values. He then measured directly the heat given off by a dog placed in a calorimeter and found that for fat and carbohydrate the heat production was the same within experimental error whether combustion occurred within or outside the body. For protein the heat production was less when metabolized (*physiological heat value*) than when burned in a calorimeter. The reason for this is that unlike carbohydrate and fat, which are completely oxidized in the body to CO_2 and water, the combustion of protein by the tissues is incomplete. The carbon part of the molecule is burned but the nitrogen is excreted in the urine as urea and other nitrogenous compounds which possess a certain energy combustion of foodstuffs in the various tissues of the body.

¹ Lavoisier, however, believed that the heat was produced through the oxidation of carbon and hydrogen in the lungs. Not until some years later was it shown that heat production was the result of the

value Burned outside the body, protein yields 5.3 Calories as compared with its physiological heat value of 4.1 The following are the quantities of heat generated by the metabolism of 1 gram of the respective foodstuffs

1 gram of fat	yields 9.3 Calories
1 gram of carbohydrate	yields 4.1 Calories
1 gram of protein	yields 4.1 Calories

The caloric values of different types of carbohydrate and protein are not identical The value for starch, for example, is 4.20 and for sugar 3.96 The value is higher for animal than for vegetable protein There is also a small energy loss during digestion For these reasons the respective values for fat, carbohydrate and protein in a mixed human diet are usually taken at 9.0, 4.0 and 4.0 respectively

(2) The energy balance

Rubner placed a full-grown dog in a calorimeter (p. 611) in which the heat production could be measured directly He fed the animal measured amounts of food for which the actual heat values had been calculated The heat generated by the animal during the experiment, which extended in most cases over several days, when added to the heat equivalent of the excreta (urine and feces) was found to correspond, within about 1 per cent, to the calculated heat value of the ingested food In other words, the energy intake and output balanced and the application to the body of the law of the conservation of energy was demonstrated (See table 43)

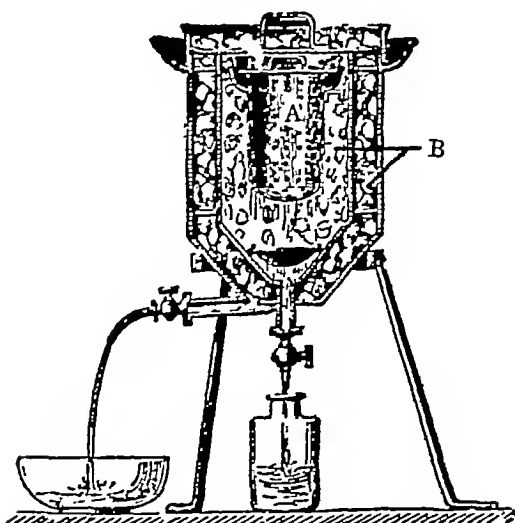


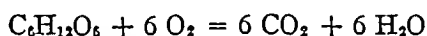
FIG. 45.1 Ice calorimeter of Lavoisier and Laplace (After Luciani) A, chamber for animal, B, two concentric chambers filled with ice

Corresponding results were obtained by Atwater for man Rubner also demonstrated that the heat production of an animal as measured directly (direct calorimetry) agrees within 1 per cent of that calculated indirectly from the respiratory exchanges (indirect calorimetry)—a brilliant confirmation of the conclusion arrived at by Lavoisier 100 years earlier

(3) The respiratory quotient

Since the relative amounts of oxygen and carbon contained in the molecules of the three foodstuffs differ, the relative volumes of oxygen consumed and of carbon dioxide produced during the metabolism of each type of food also varies The ratio $\frac{\text{vol CO}_2 \text{ expired}}{\text{vol O}_2 \text{ consumed}}$ is called the respiratory quotient or, briefly, the R. Q.

The following equation represents the oxidation of glucose



It is clear from this equation that no oxygen is required from a source outside the food itself for the oxidation of the hydrogen in the glucose molecule, and that for each molecule of O_2 absorbed a molecule of CO_2 is produced It will also be recalled that according to the law of Avogadro quantities of any two gases containing the same number of molecules and under identical conditions of temperature and pressure have equal volumes In the metabolism of 100 grams of glucose, 75 liters of oxygen are required from the outside for its complete combustion, i.e., for the oxidation of the carbon, 75 liters of carbon dioxide are therefore produced The respiratory quotient, then is $75/75 = 1.0$ For the combustion of 100 grams of fat, (e.g. triolein $\text{C}_{57}\text{H}_{104}\text{O}_2$), which is rich in carbon and hydrogen but relatively poor in oxygen, 200 liters of the latter gas are required, 142 liters of carbon dioxide and about 110 grams of water are produced The R.Q. is therefore $142/200 = 0.71$ The R.Q. of protein is 0.80 and of alcohol 0.67 The value of the respiratory quotient is taken as an indication of the type of food being metabolized It gives no quantitative estimation of the metabolism An R.Q. around unity is taken to indicate that the material being used is chiefly carbohydrate, one around 0.70 indicates that it is mainly fat²

² Cathcart and Markowitz point out, however, that probably too much reliance has been placed upon the value of the respiratory quotient as an unequivocal

TABLE 43*
Comparison of estimated heat from metabolism of food with heat actually produced

FOOD	NUM BER OF DAYS	HEAT CALCU- LATED	HEAT DIRECTLY DETER- MINED	DIFFERENCE IN PER CENTAGE
		cal	cal	
Starvation	5	1296 3	1305 2	-1 42
	2	1091 2	1056 6	
Fat	5	1510 1	1498 3	-0 97
Meat and fat	8	2492 4	2488 0	
	12	3985 4	3058 4	
Meat	6	2249 8	2276 9	-0 42
	7	4780 8	4769 3	

* From Lusk after Rubner

On an ordinary mixed diet the respiratory quotient is about 0.85 and in the postabsorptive state (p. 617) about 0.82. In the formation of fat from carbohydrate, as in the fattening of farm animals, an oxygen-rich substance is being converted into one poor in oxygen. Oxygen is liberated in the conversion and less, in consequence, is taken in from the outside for general oxidative processes. The respiratory quotient may therefore rise above unity (up to 1.4). A very low quotient, as is seen in the hibernating animals, is supposed by some to indicate the reverse process, i.e., the conversion of fat to carbohydrate. In the hibernating marmot the R.Q. is between 0.6 and 0.7.

(4) *Thermal equivalents*

As mentioned on page 606 the heat production bears a relation to the oxygen consumed and the carbon dioxide eliminated. Consequently, if one knows the quantity of either of these respiratory gases consumed or produced, respectively, during a given period, the heat production of the body during that time can be calculated. The heat production for a given quantity of these gases (*thermal equivalent*) varies, however, with the type of food undergoing combustion. For instance, when carbohydrate is burned the consumption of a liter of oxygen causes a greater evolution of heat (0.047

criterion of the type of foodstuff undergoing metabolism. The R.Q. in a given instance is undoubtedly a resultant of several different metabolic processes—syntheses and interconversions as well as combustion.

Cal.) than when fat is burned (4.686 Cal.) This is so because a larger part of the oxygen required for the complete combustion of the former substance is, as mentioned above, contained within its own molecule. The value for protein is 4.485 Calories. The heat equivalent of CO₂ varies much more than does that of oxygen, for example, a liter of CO₂ formed in the combustion of carbohydrate represents the evolution of 5.047 Calories, whereas, the heat equivalent of this volume of CO₂, when the fuel is fat, is 6.629 Calories.

It is therefore more usual to calculate the heat production from the oxygen consumption, but even so, the food mixture undergoing combustion, as indicated by the respiratory quotient, must be taken into account. The heat or caloric values (*thermal equivalent*) of a liter of oxygen at different respiratory quotients is given in table 44 compiled by Zuntz and Schumburg (as modified by Lusk).

In table 44 the percentages of fat and carbohydrate undergoing combustion have been calculated for respiratory quotients ranging from 0.707 to 1.0. These so-called *non protein respiratory quotients* were obtained by determining the total oxygen consumption and carbon dioxide produced and then subtracting the volumes of these gases exchanged in the catabolism of protein. The quantity of protein undergoing catabolism is obtained from the urinary nitrogen, each gram of the latter being equivalent to 6.25 grams of protein. In precise experiments upon heat production, the calories produced from the catabolism of protein as well as those derived from fat and carbohydrate would require to be determined. For example, a subject produces per hour 13.50 liters of carbon dioxide, consumes 16.00 liters of oxygen and excretes 0.5 gram of nitrogen in the urine. Now, each gram of urinary nitrogen represents the production of 4.76 liters of carbon dioxide and the consumption of 5.94 liters of oxygen.

Therefore
the CO₂ produced by the subject in the catabolism of protein is $0.5 \times 4.76 = 2.38$ liters
the O₂ consumed in the catabolism of protein is $0.5 \times 5.94 = 2.97$ liters

Then
the non protein CO₂ production is $13.50 - 2.38 = 11.12$ liters
the non protein O₂ consumption is $16.00 - 2.97 = 13.03$ liters

the non-protein re-
spiratory quotient
is

$$\frac{11 \ 12}{13 \ 03} = 0 \ 85$$

It will be seen from table 39 that at this R Q the caloric equivalent of a liter of oxygen is 4 862. The heat produced by the combustion of non-protein materials is therefore, $13 \ 03 \times 4 \ 862 = 63 \ 3$ Calories of which 50 7 per cent is derived from carbohydrate and 49 3 per cent from fat.

The heat production due to protein is $2 \ 97 \times 4 \ 485 = 13 \ 3$ Calories. (The caloric equivalent of each gram of urinary N is 26 51, so, the heat production due to protein may also be obtained approximately by multiplying the figure for the urinary nitrogen (0 5) by 26 51.)

The total heat production per hour is therefore $63 \ 3 + 13 \ 3 = 76 \ 6$ Calories. Of this, 42 per cent is derived from carbohydrate, 41 per cent from fat and 17 per cent from protein.

In ordinary determinations of the basal metabolic rate (p 617) urinary nitrogen is not measured and the foregoing calculations are not undertaken, only a slight error is involved if the R Q is assumed to be 0 82 and the heat production taken directly from table 44.

(5) *The isodynamic law*

It was demonstrated by Rubner that just as the production of heat by a stove may be maintained at a constant level by burning different types of fuel, so in the generation of animal heat the different foodstuffs may replace one another in the diet in accordance with their heat producing values. Thus —

100 grams of fat

232 grams of starch

234 grams of cane sugar

243 grams of dried meat

produce equivalent amounts of heat when ingested, i.e., they are isodynamic.

(6) *Heat production in relation to surface area*

The heat produced by an individual at rest is proportional to the surface area of his body. Thus a fasting adult man and a starving dog, though the surface area of each and the total heat production were widely different, were shown by Rubner to give out in 24 hours closely similar amounts of heat *per square meter of body surface*, namely, 1134 and 1112 Calories respectively. A

TABLE 44

(After Zuntz and Schumberg, modified by Lusk*)

NO. PROTEIN RESPIRATORY QUOTIENT	CALORIES PER LITER O ₂	CALORIES DERIVED FROM	
		Carbohydrate	Fat
		<i>per cent</i>	<i>per cent</i>
0 707	4 686	0	100
0 71	4 690	1 10	98 9
0 72	4 702	4 76	95 2
0 73	4 714	8 40	91 6
0 74	4 727	12 0	88 0
0 75	4 739	15 6	84 4
0 76	4 751	19 2	80 8
0 77	4 764	22 8	77 2
0 78	4 776	26 3	73 7
0 79	4 788	29 9	70 1
0 80	4 801	33 4	66 6
0 81	4 813	36 9	63 1
0 82	4 825	40 3	59 7
0 83	4 838	43 8	56 2
0 84	4 850	47 2	52 8
0 85	4 862	50 7	49 3
0 86	4 875	54 1	45 9
0 87	4 887	57 5	42 5
0 88	4 899	60 8	39 2
0 89	4 911	64 2	35 8
0 90	4 924	67 5	32 5
0 91	4 936	70 8	29 2
0 92	4 948	74 1	25 9
0 93	4 961	77 4	22 6
0 94	4 973	80 7	19 3
0 95	4 985	84 0	16 0
0 96	4 998	87 2	12 8
0 97	5 010	90 4	9 58
0 98	5 022	93 6	6 37
0 99	5 035	96 8	3 18
1 00	5 047	100 0	0

* This table has been further modified by Cathcart and Cuthbertson, see J. Physiol., 1931, 72, 349.

small animal, e.g., a mouse, therefore, since its surface area is greater in proportion to its mass, and since it generates the same amount of heat per unit of body surface, must obviously generate more heat per unit of body weight than a larger animal. The heat is produced in the tissues (muscles, liver, etc.), consequently these, in the case of the smaller animal, must be the seat of a much more active metabolism (see table 45).

(7) *Physiological conditions which stimulate metabolism*

The heat production of the body is increased by (a) muscular work, (b) food, (c) a fall in

TABLE 45

Showing relation of heat production per kilogram and per square meter of body surface in animals of different sizes

	WEIGHT	CALORIES	
		Per kilogram	Per square meter surface
	kg		
Pig	128 0	19 1	1,078
Man	64 3	32 1	1,042
Dog	15 2	51 5	1,039
Goose	3 5	66 7	967
Fowl	2 0	71 0	947
Mouse	0 018	654 0	1,188

The relatively low figures for the heat production of birds shown in the last column is due to their bodies containing a high proportion of osseous tissue which has an extremely low metabolism.

environmental temperature, or (d) a rise in body temperature (fever) These influences will be considered later

CALORIMETRIC METHODS

THE BOMB CALORIMETER

The bomb calorimeter is employed for the determination of the potential energy of food. The food is dried and a weighed quantity ignited, by means of an electric fuse, in a hollow steel container lined with platinum and filled with pure oxygen (fig 45.2). The heat evolved (energy of the food converted to heat) is absorbed by a known quantity of water in which the container is immersed. The quantity of water in kilograms (plus the water equivalent of the apparatus) multiplied by the number of degrees centigrade through which its temperature has been raised gives, in Calories, the quantity of heat generated.*

CALORIMETRIC MEASUREMENTS IN ANIMALS AND MAN

An animal's energy is derived from the food which is to the body what fuel is to a furnace or

* Another type of calorimeter called the *oxy-calorimeter* has been devised by Benedict. The caloric value of the food is determined by burning a sample in a special combustion chamber and measuring the volume of oxygen used. Knowing the thermal equivalent of a liter of oxygen used in the combustion of the three food stuffs (p. 609) the heat value of the particular food sample is readily calculated. This method is really an adaptation of the method of indirect calorimetry used in the clinic and described on page 612.

machine. We have seen that the law of conservation of energy holds true for the animal body—that in a healthy animal which is maintaining a constant weight the intake and output of energy are equal. The food undergoes combustion in the tissues, its carbon being oxidized to carbon dioxide, its hydrogen to water and its potential energy converted into other forms of energy—mechanical, electrical, chemical and thermal. Thus the various processes of life are sustained. In a growing animal or in an animal during fattening the energy of the food is in part stored as new-formed tissue.

In the resting body all the energy liberated from the food ultimately appears as heat. A heat unit has therefore been chosen as the most convenient one for measuring and expressing the energy exchanges of the body. This unit is the *large calorie* (Calorie or Cal), i.e., the quantity of heat required to raise the temperature of a kilogram of water 1 degree centigrade (actually from 15° to 16°C). The Calorie used in physiological determinations is therefore 1000 times the *small calorie* (calorie or cal) used in physical measurements.

Calorimetric methods are of two main types: (a) *direct* which is the same in principle as that described above for the bomb calorimeter, and (b) *indirect* in which the heat production is calculated from the respiratory exchanges.

Direct calorimetry

The elaborate nature of the apparatus required for this method precludes its use, especially for the human subject, in any ordinary laboratory or clinic. There are indeed only a very few in existence.

The apparatus consists of an insulated chamber large enough to accommodate an animal or a man. The heat eliminated from the body is absorbed by water circulating in coils of copper pipes. The temperature within the chamber is kept constant and the temperature of the subject is taken from time to time to ensure that heat is not being retained by the body. The heat generated and radiated by the subject is absorbed by the circulating water. The temperature of this as it enters and leaves the chamber being known, and also its quantity (kilograms), the heat production in Calories can be determined. To this must be added the latent heat of the water vapor given off by the lungs and skin. This, which amounts to about one-quarter of the total heat production, is calculated from the weight of the water in grams removed from the air by an absorber

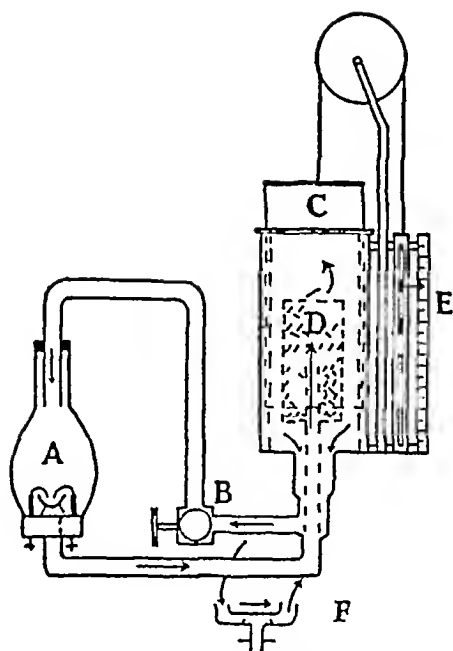


FIG 45.2 Diagram of oxy-calorimeter for determining energy values of foods, etc (A) Combustion chamber, crucible and electrical connections (B) Rotary blower (C) Spirometer bell with oxygen in water seal (D) Soda lime container (E) Scale in millimeters for measuring the oxygen used (F) Mouthpiece with valves for BMR estimation (enlarged) Principle of the Benedict-Roth apparatus for determining BMR is seen by replacing A and B by I' (After Cruickshank, Livingston and Co, Pub, Edin)

(sulphuric acid) multiplied by the factor 0.59 Calorie which is the latent heat of 1 gram of vaporized water at 20°C. A calorimeter of this type is usually combined with apparatus for determining the heat production by indirect calorimetry as described below, the instrument being then referred to as a respiration calorimeter (fig 45.3)

Indirect calorimetry determination of the heat production from the respiratory exchanges

Experiments by several observers upon man and animals have shown that the results of indirect calorimetry agree within less than 1 per cent with those obtained by the direct method. Two forms of apparatus—the *closed-circuit* and the *open-circuit* or *air-current* types—are employed for indirect calorimetry. In the first mentioned method the subject rebreathes the air contained in a closed system, the carbon dioxide eliminated by the subject is removed by soda-lime and weighed, a measured volume of oxygen is supplied to replenish that which has been absorbed. In the open-circuit type, the subject inspires room air and expires into some form of container the entire volume of expired air is measured and a sample analyzed for its carbon dioxide and oxygen

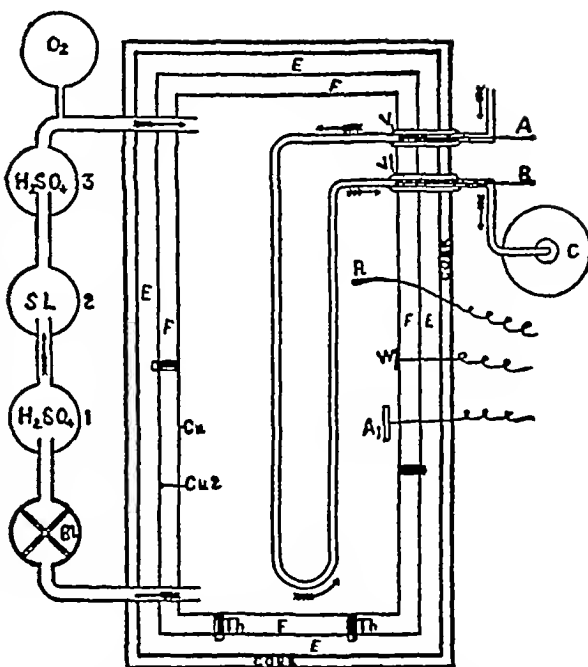


FIG 45.3 Schematic diagram of the Atwater-Rosa-Benedict respiration calorimeter Ventilating system. O_2 , oxygen introduced as consumed by subject, 3, H_2SO_4 to catch moisture given off by soda lime, 2, soda lime to remove CO_2 , 1, H_2SO_4 to remove moisture given off by patient, B1, blower to keep air in circulation. Indirect calorimetry. Increase in weight of H_2SO_4 (1) = water elimination of subject, increase in weight of soda lime (2) + increase in weight of H_2SO_4 (3) = CO_2 elimination, decrease in weight of oxygen tank = oxygen consumption of subject. Heat-absorbing system A_1 , thermometer to record temperature of ingoing water, B, thermometer to record temperature of outgoing water, V, vacuum jacket, C, tank for weighing water which has passed through calorimeter each hour, W, thermometer for measuring temperature of wall, A_1 , thermometer for measuring temperature of the air, R, rectal thermometer for measuring temperature of subject. Direct calorimetry. Average difference of A and B \times liters of water + (gm water eliminated \times 0.586) \pm (change in temperature of wall \times hydrothermal equivalent of box) \pm (change of temperature of body \times hydrothermal equivalent of body) = total calories produced. Th, thermocouple, Cu, inner copper wall, Cu₂, outer copper wall, E, F, dead air-spaces (After Lusk, *The Science of Nutrition*)

percentages. Among the closed-circuit types of apparatus are those of Regnault and Reiset, and of Benedict and associates. The Douglas bag and Tissot methods are of the open-circuit type. In Haldane's method for small animals, though it is of the open-circuit or air-current type, the carbon dioxide is absorbed and weighed.

THE CLOSED-CIRCUIT METHODS

In the Regnault-Reiset type of apparatus, the air is circulated through a closed system of which a chamber, large enough to accommodate the subject, forms a part.

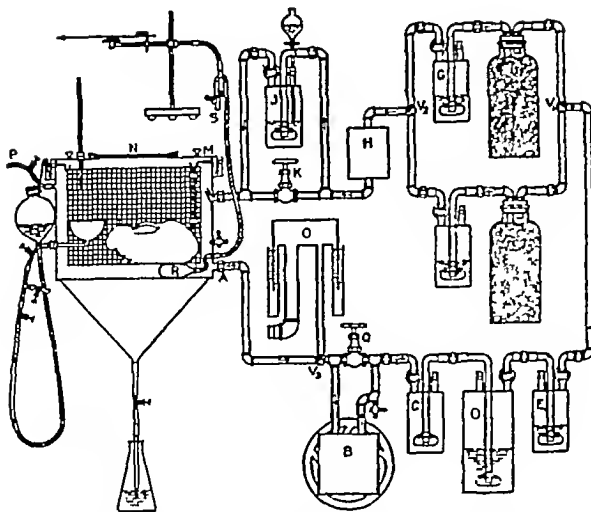


FIG 45.4 Schematic outline of respiration apparatus for small animals. The air leaves the chamber A and after passing through the rotary blower B, which keeps the ventilating current in motion, is forced through the glass vessel C, which serves as a safety trap. The air then passes through the bottles D and E containing sulphuric acid to remove the moisture. The air, now water free, but containing the CO_2 produced by the animal and lacking the O_2 which the animal has consumed, passes into the 2 way valve V_1 , where it may be deflected into the bottle F containing soda lime for the absorption of CO_2 . The moisture gained from the soda lime is absorbed by sulphuric acid in bottle G. The air then passes through a second 2 way tap V_2 to H containing dry sodium bicarbonate which removes the traces of acid vapor taken up by the air in passing through bottle G. J is a glass vessel containing water which supplies sufficient moisture to the air for the comfort of the animal. K is a by-pass valve. The chamber is constructed of copper and has a cover with a water seal M. N is a glass plate through which the animal can be observed. O is a spirometer attached to the system on the intake side of the rotary blower B. (From F. G. Benedict.)

The air in the system is rebreathed repeatedly, carbon dioxide and water vapor being removed and oxygen supplied to replace that consumed. The air upon leaving the chamber enters first an absorbing vessel containing sulphuric acid which removes the moisture, then through a container filled with moist soda-lime which removes the carbon dioxide, and finally through a second sulphuric acid container which abstracts the water gained from the soda lime. The dried and carbon dioxide free air is then returned to the chamber. The quantity of carbon dioxide eliminated by the animal is given by the difference between the weights of the soda lime container at the beginning and end of the experiment. Oxygen is run into the system from a cylinder and measured by means of a gas meter or by weighing the cylinder at the beginning and end of the experiment. The air in the chamber is analyzed at the end of the observation in order to ensure that no change in its composition has occurred. Only a few institutions on this continent, such as the Russell Sage Institute in New York and the Nutrition Laboratory at Washington, possess an apparatus of this type suitable for metabolic studies upon man. Such an apparatus was first constructed in America by Atwater and Rosa. It is usually combined with an apparatus of the direct type (fig 45.3). The construction of a closed circuit type of apparatus for laboratory

animals is not, however, such a difficult matter (fig 45.4).

In other closed-circuit methods, such as the one described below, the subject is not enclosed within a chamber, he simply breathes in and out of the apparatus through a connecting tube.

Clinical types of closed circuit apparatus. In the earlier clinical types the subject was connected by means of a mouthpiece and flexible tubing to the closed system and both carbon dioxide elimination and oxygen usage were determined. In the type most commonly used today—the Benedict-Roth apparatus (figs 45.5 and 45.6)—the heat production is calculated from the oxygen consumption alone. In order to purify the air the carbon dioxide is absorbed by soda-lime but the amount of this gas eliminated is not measured. In determining the basal metabolism (p 617) the subject lies upon a couch and breathes in and out of the instrument through a mouthpiece and two wide bore tubes (inspiratory and expiratory) provided with valves. The nose is clipped. The main part of the instrument consists of a bell-type spirometer. This is a hollow double-walled cylinder.

drical vessel. In the narrow space between the two walls fits a second inverted hollow cylindrical vessel or bell. The bell is counterpoised so that it rides easily up or down in the annular space between the two walls. This space contains water which acts as a seal. At the commencement of the experiment, sufficient oxygen is admitted from an oxygen cylinder to raise a pointer on the spirometer bell to the zero mark upon an adjoined scale which has been calibrated to oxygen volumes. The breathing of the subject through the inspiratory and expiratory tubes keeps the air circulating freely through the system. As oxygen is consumed the spirometer falls and from the difference in the levels of the pointer at the beginning and end of the experiment the oxygen usage is calculated. The volume of oxygen used, dry and reduced to standard temperature and pressure (see p 616) is then calculated. The heat production is found by reference to the table of respiratory quotients given on page 609. The average RQ in the postabsorptive state (see p 617) is 0.82. Since the RQ of the subject is not determined by the foregoing method this value (0.82) is assumed. It will be seen from table 44 that at this RQ the caloric equivalent of 1 liter of oxygen is 4.825.

OPEN-CIRCUIT OR AIR-CURRENT METHODS

In the *Haldane type* of calorimeter (which is suitable only for small animals, mice, rats or rabbits) a current of air is drawn through the system (fig 45 7). Carbon dioxide and water are removed from the air before it enters the chamber and again after its exit therefrom. The carbon dioxide absorber on the outgoing current of air is weighed at the commencement and end of the experiment as in the Regnault-Reiset method, in order to obtain the quantity of carbon dioxide eliminated. The system with the exception of the first pair of absorbers and including the animal is then weighed. Since only oxygen has entered this part of the system (the air being CO_2 -free and dry) the gain in weight during the experiment gives the quantity of oxygen consumed by the animal.

The Douglas bag and Tissot methods (gasometric methods) In these methods the subject inspires atmospheric air and expires into a bag (Douglas method) or into a large bell-type spirometer (Tissot method). At the end of the experiment the total volume of expired air is measured and samples are analyzed for carbon dioxide and oxygen (figs 45 8 and 45 9).

Gas analysis The sample of expired air is drawn



FIG 45 5 Benedict-Roth apparatus (courtesy of Warren E Collins, Inc, Boston)

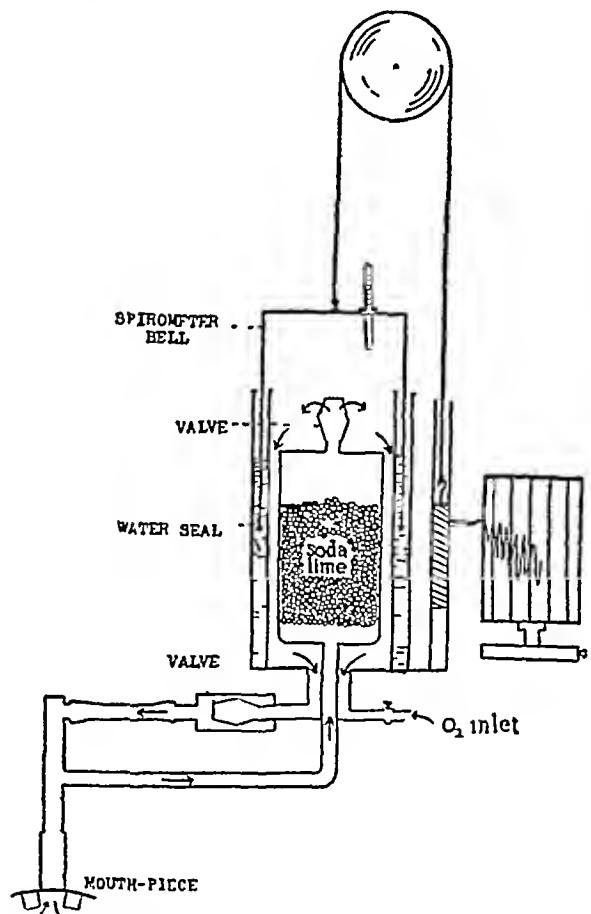


FIG 45 6 Sectional view of Benedict-Roth closed-circuit respiration apparatus

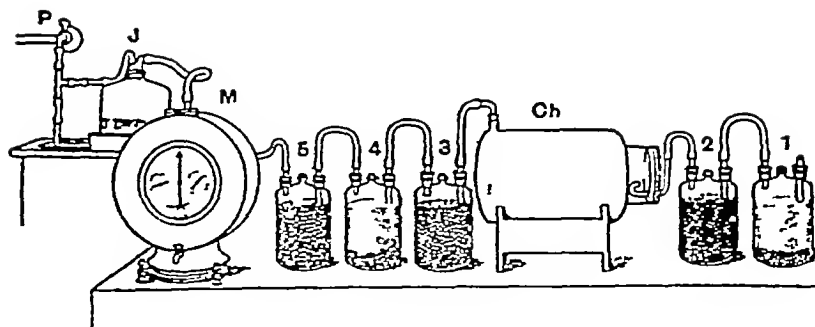


FIG 45.7 Haldane's respiration apparatus. 1 and 4, soda lime, 2, 3 and 5, pumice stone soaked in sulphuric acid, Ch, animal chamber, M, meter J is an inverted bell jar standing in a trough of water, it serves to prevent sudden excess of negative pressure and to indicate the pressure actually employed P, pump (After Haldane.)

into the graduated burette of a Haldane gas analysis apparatus, saturated with water vapor and its volume measured (fig 45.10). It is then passed back and forth through the bulb containing potassium hydroxide solution which absorbs the carbon dioxide. It is then measured again. The difference between the two measurements gives the volume of carbon dioxide in the sample. Next, the oxygen is removed by passing the sample through a solution of potassium pyrogallate. The sample is measured a third time and the shrinkage in volume as shown by the difference between

the second and third readings gives the quantity of oxygen absorbed. From the data so obtained the percentages of carbon dioxide and oxygen in the sample of air are calculated.

Calculation of results The following illustrates the steps in an actual metabolism experiment, using the Douglas or Tissot method.

Period of observation 10 minutes, barometer 751.5 mm Hg

Volume of expired air as determined by passing the expired air, collected in a bag, through a meter (Douglas method) or as indicated by spirometer (Tissot method) = 70 liters.

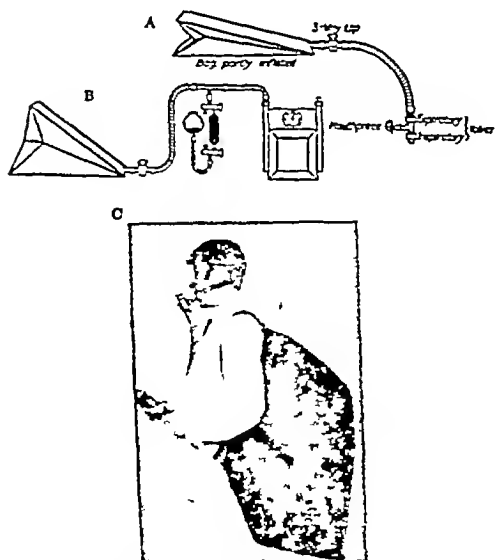


FIG 45.8 Showing A, Douglas bag and tubing, B, Douglas bag with sampling bulb attached and gas meter for measuring the total volume of expired air (after Douglas and Priestley), C, subject equipped with Douglas bag apparatus during running or other types of muscular exercise. (After Hill.)

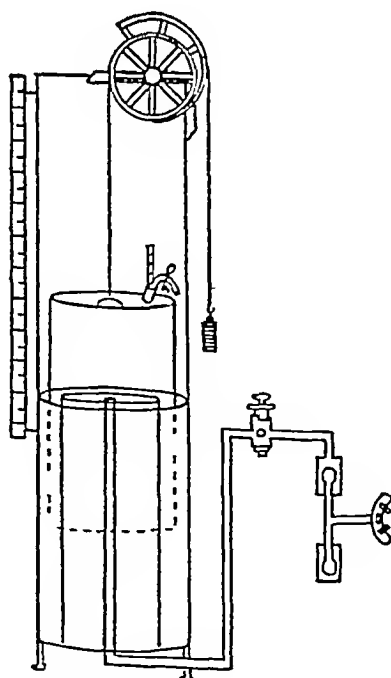


FIG 45.9 Diagram of Tissot spirometer

TABLE 46
Table for reduction to 0°C and 760 mm Hg and dry of 1 liter of air saturated with humidity, from 10° to 25°C, and 740 to 780 mm (29.13 inches to 30.71 inches) of mercury
(intermediate values may be obtained by interpolation)

TEMPERATURE	29 13 740	29 31 742	29 29 744	29 37 746	29 45 748	29 53 750	29 60 752	29 68 754	29 76 756	29 84 758	29 92 760	30 00 762	30 08 764	30 16 766	30 24 768	30 31 770	30 39 772	30 47 774	30 55 776	30 63 778	30 71 780	TEMPERATURE
10	927 7	930 2	932 6	935 1	937 6	940 4	942 9	945 4	947 9	950 5	953 0	955 6	958 0	960 6	963 1	965 7	968 3	970 8	973 3	975 9	978 4	10
11	923 6	926 1	928 5	931 0	933 5	936 3	938 8	941 3	943 8	946 4	948 9	951 5	953 9	956 5	959 0	961 6	964 1	966 6	969 1	971 9	974 2	11
12	919 5	921 8	924 2	926 7	929 3	931 8	934 3	936 8	939 4	942 0	944 4	947 0	949 4	951 9	954 4	957 0	959 5	962 0	964 5	967 1	969 6	12
13	915 4	918 0	920 4	922 9	925 4	928 0	930 4	932 9	935 5	938 1	940 5	943 1	945 5	948 1	950 6	953 1	955 6	958 1	960 6	963 2	965 7	13
14	911 3	913 9	916 3	918 8	921 3	923 8	926 2	928 8	931 3	933 9	936 2	938 9	941 3	943 8	946 3	948 8	951 3	953 8	956 3	958 8	961 3	14
15	907 1	909 7	912 1	914 6	917 1	919 6	922 0	924 5	927 1	929 7	932 0	934 6	937 0	939 5	942 0	944 4	947 0	949 6	952 0	954 5	957 0	15
16	902 9	905 5	907 9	910 4	912 9	915 4	917 8	920 3	922 8	925 4	927 8	930 4	932 8	935 2	937 7	940 1	942 6	945 2	947 6	950 1	952 6	16
17	898 7	901 3	903 7	906 2	908 7	911 1	913 5	916 0	918 5	921 1	923 5	926 0	928 5	930 9	933 4	935 8	938 3	940 9	943 3	945 8	948 3	17
18	894 5	897 1	899 5	902 0	904 5	906 8	909 2	911 8	914 2	916 8	919 2	921 7	924 2	926 6	929 1	931 5	933 9	936 5	938 9	941 4	943 9	18
19	890 2	892 7	895 1	897 6	900 1	902 5	904 9	907 4	909 9	912 5	914 8	917 2	919 7	922 2	924 7	927 1	929 5	932 0	934 4	936 9	939 4	19
20	885 9	888 4	890 8	893 3	895 8	898 1	900 5	902 9	905 3	907 7	910 4	912 8	915 2	917 7	920 2	922 6	925 0	927 5	930 3	932 5	935 0	20
21	881 8	884 3	886 7	889 2	891 7	894 0	896 4	898 9	901 3	903 9	906 2	908 6	911 1	913 5	916 0	918 4	920 8	923 3	925 7	928 2	930 7	21
22	877 1	879 5	881 9	884 4	886 9	889 0	891 4	894 1	896 6	899 2	901 4	903 8	906 3	908 7	911 2	913 6	916 0	918 4	920 9	923 4	926 0	22
23	872 6	875 0	877 4	879 9	882 4	884 7	887 1	889 5	892 0	894 6	896 9	899 2	901 7	904 1	906 6	909 0	911 4	913 8	916 3	918 8	921 3	23
24	868 1	870 6	873 0	875 5	878 0	880 1	882 5	885 0	887 5	890 1	892 3	894 6	897 1	899 5	902 0	904 4	906 8	909 2	911 6	914 0	916 5	24
25	863 5	865 9	868 3	870 8	873 3	875 7	878 1	880 5	882 9	885 5	887 9	890 1	892 6	895 0	897 4	899 8	902 1	904 5	906 9	909 3	911 7	25

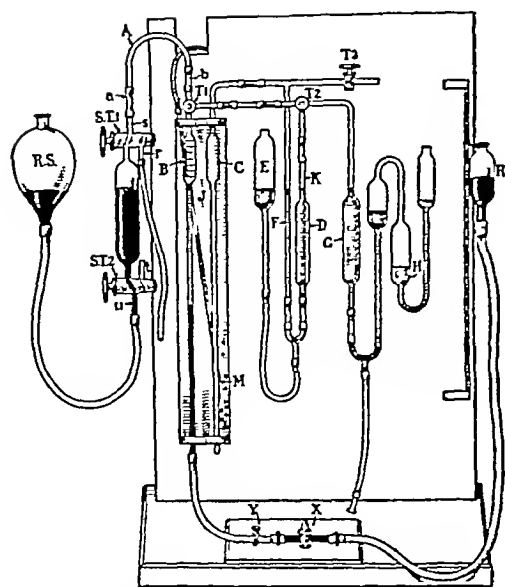


FIG 45 10 Haldane gas analysis apparatus (small pattern) showing a gas sampler in position. A, glass tube connected to gas sampler, B, gas burette, C, control tube, D, caustic soda absorption pipette, E, caustic soda reservoir, F, caustic pressure tube, G, pyrogallol absorption pipette, H, caustic soda seal, J, waterjacket, M, aeration tube, R, mercury reservoir, X, one way tap to control movements of mercury, Y, screw clip for fine adjustments of pressure in K after the tap X has been closed, R.S., mercury reservoir for gas sampler, S.T. 1 and S.T. 2, taps of gas sampler with double ports, r and t, side tubes by means of which, when S.T. 1 or S.T. 2 are placed in position, the "dead spaces" of these taps and of the tube A can be filled with mercury, a, rubber connection, any air bubble in A may be expelled here as described in text, F, pressure tube for control, K, pressure tube for burette. Note T3 is usually a three way tap, placed at the junction of the vertical and horizontal tubes (From Lamb, *An Introduction to Human Experimental Physiology*)

Temperature of expired air (in meter or spirometer) 20°C.

The volume of the gases must be reduced to standard conditions, namely, to 760 mm Hg pressure and to 273° absolute temperature (i.e., to 0°C), and dry (i.e., the pressure of water vapor must be deducted from the barometer reading) The pressure of water vapor at 20°C is 17.5 mm So the corrected reading is,—

$$751.5 - 17.5 = 734.0 \text{ mm}$$

The volume of the expired air at standard pressure and temperature (S.P.T.) and dry is therefore

$$70 \times \frac{734}{760} \times \frac{273}{273 + 20}$$

or

$$70 \times 0.8993 = 62.95 \text{ liters during the period of observation (10 minutes) or } 6,295 \text{ cc. per minute}$$

In practice these detailed calculations are avoided by reference to table 46 which gives the required factor by which the observed volume is multiplied in order to reduce it to standard conditions and dry

Results of gas analyses

Expired air	Inspired air
CO = 3.50 per cent	CO = 0.04 per cent
O ₂ = 16.90 per cent	O ₂ = 20.93 per cent
N = 79.60 per cent	N ₂ = 79.03 per cent

Since the O₂ percentages in expired and inspired airs are 16.90 and 20.93 respectively it might be thought that 20.93 — 16.90 would give the percentage of O₂ absorbed. It will be noticed, however, that the percentage of N₂ is higher in the expired than in the inspired air. Nitrogen is inert insofar as respiration is concerned, being neither produced nor retained in the body, i.e., its absolute amount is not altered. Therefore, its greater proportion in the expired air can only mean that the volume of the inspired air (which of course was not measured) must have been greater than that of the dry expired air. So then, the volume of O₂ inspired must also have been greater than appeared from the analysis of the expired air. The cause of the discrepancy is that part of the absorbed oxygen has combined with hydrogen and in other ways, and so has not appeared as CO. The extent to which the O₂ in the inspired air exceeds that shown by the analysis of the expired air is proportional to the increased percentage of N in the latter. Instead of the inspired air having contained 20.93 volumes of O for every 100 volumes of air expired, it must have contained

$$20.93 \times \frac{79.60}{79.03}, \text{ or } 0.265 \text{ (a constant factor)} \times 79.60 = 21.09 \text{ volumes}$$

Therefore the O₂ absorption is

$$\frac{21.09 - 16.90}{100} \times 6295 = 264 \text{ c.c. per minute}$$

The calculation is abbreviated by the use of table 47

The quantity of CO produced may be calculated without correction since its percentage in the inspired air is negligible. Hence

$$\frac{3.50 - 0.04}{100} \times 6295 = 218 \text{ cc per minute}$$

The respiratory quotient is therefore

$$\frac{\text{Vol CO}_2 \text{ expired } 218}{\text{Vol O}_2 \text{ consumed } 264} = 82$$

The caloric values of O_2 and CO_2 are given in, or can be calculated from table 44, p 609. The heat production may be calculated from either of these values. For example, when the R:Q is 82 the caloric value of 1000 cc of O_2 is 4825. Therefore, 264 cc of O_2 represents a heat production of $4825 \times (264/1000) = 127$ cal per minute or 7620 cal per hour.

THE BASAL METABOLIC RATE (BMR)

This is the term applied to the heat production of a subject who though awake is as nearly as possible at complete *muscular* and *mental rest*, and is in the *postabsorptive* state (i.e., from 12 to 14 hours after a light meal when, it is assumed, the digestive processes are quiescent). The room temperature should be $20^\circ C$. For example, the prospective subject of a basal rate determination is directed to refrain from undue muscular exertion or fatiguing effort of any kind for 24 hours previously. A light meal with the minimum amount of protein is taken not later than 7 o'clock the night before the test which is undertaken at about 9 o'clock in the morning. For a period of 30 minutes or so before, as well as during the test the subject should be lying down comfortably in a room with subdued lighting.

The apparatus most commonly employed for the determination of the basal metabolic rate in the clinic is the instrument of Benedict and Roth (p 612). Tissot's method or the bag method of Douglas is also sometimes used.

BASAL METABOLIC RATE STANDARDS

We have seen that the metabolism is proportional to the body's surface area rather than to its weight (p 609). For example, of two men of about the same age, one large and the other small, though the former has the greater total heat production, the heat production per unit of body surface is of practically the same value in both subjects. Consequently, since the surface area of the larger man is, in proportion to his mass, less than that of the smaller, his heat production per unit of body weight must be less. In figure 45 11, A, are shown two wooden blocks. The volume and weight of the larger is some three times greater than that of the smaller, but its surface area is only twice as great. If the metabolic rate were accurately related to weight rather than to surface area, then of two similarly shaped animals differing in size to the same extent as the two blocks the metabolism of the larger would be

TABLE 47
Volumes of oxygen in inspired air corresponding to 100 volumes of expired air with different percentages of nitrogen

PER CENT NITROGEN IN EXPIRED AIR	VOLUMES OF OXYGEN IN INSPIRED AIR	PER CENT NITROGEN IN EXPIRED AIR	VOLUMES OF OXYGEN IN INSPIRED AIR
78.5	20.80	79.6	21.09
78.6	20.83	79.7	21.12
78.7	20.85	79.8	21.14
78.8	20.88	79.9	21.17
78.9	20.91	80.0	21.20
79.0	20.93	80.1	21.22
79.1	20.96	80.2	21.25
79.2	20.99	80.3	21.28
79.3	21.01	80.4	21.30
79.4	21.04	80.5	21.33
79.5	21.07		

three times greater than that of the smaller, it is found, however, to be only about twice as great.

Also two objects of the same volume and weight will have different surface areas if they differ in shape. In figure 45 11, B, the board and block are identical in volume and weight but the surface area of the former is more than double that of the latter. The nearer an object approaches in shape

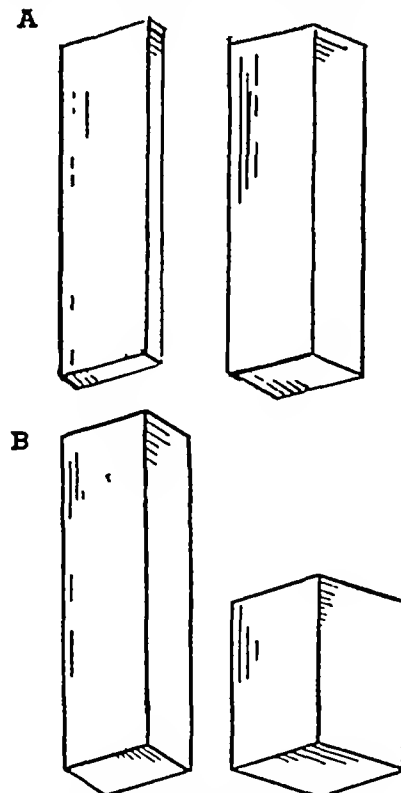


FIG 45 11 Description in text.

TABLE 48
(Relation of height and weight to surface area after Du Bois)

HEIGHT IN CENTI METERS	WEIGHT IN KILOGRAMS																
	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105
200							1.84	1.91	1.97	2.03	2.09	2.15	2.21	2.26	2.31	2.36	2.41
195						1.73	1.80	1.87	1.93	1.99	2.05	2.11	2.17	2.22	2.27	2.32	2.37
190				1.56	1.63	1.70	1.77	1.84	1.90	1.96	2.02	2.08	2.13	2.18	2.23	2.28	2.33
185				1.53	1.60	1.67	1.74	1.80	1.86	1.92	1.98	2.04	2.09	2.14	2.19	2.24	2.29
180				1.49	1.57	1.64	1.71	1.77	1.83	1.89	1.95	2.00	2.05	2.10	2.15	2.20	2.25
175	1.19	1.28	1.36	1.46	1.53		1.67	1.73	1.79	1.85	1.91	1.96	2.01	2.06	2.11	2.16	2.21
170	1.17	1.26	1.34	1.43	1.50	1.57	1.63	1.69	1.75	1.81	1.86	1.91	1.96	2.01	2.06	2.11	
165	1.14	1.23	1.31	1.40	1.47	1.54	1.60	1.66	1.72	1.78	1.83	1.88	1.93	1.98	2.03	2.07	
160	1.12	1.21	1.29	1.37	1.44	1.50	1.56	1.62	1.68	1.73	1.78	1.83	1.88	1.93	1.98		
155	1.09	1.18	1.26	1.33	1.40	1.46	1.52	1.58	1.64	1.69	1.74	1.79	1.84	1.89			
150	1.06	1.15	1.23	1.30	1.36	1.42	1.48	1.54	1.60	1.65	1.70	1.75	1.80				
145	1.03	1.12	1.20	1.27	1.33	1.39	1.45	1.51	1.56	1.61	1.66	1.71					
140	1.00	1.09	1.17	1.24	1.30	1.36	1.42	1.47	1.52	1.57							
135	0.97	1.06	1.14	1.20	1.26	1.32	1.38	1.43	1.48								
130	0.95	1.04	1.11	1.17	1.23	1.29	1.35	1.40									
125	0.93	1.01	1.08	1.14	1.20	1.26	1.31	1.36									
120	0.91	0.98	1.04	1.10	1.16	1.22	1.27										

to a cube or a sphere the less is its surface area in proportion to its bulk. So, of two men of the same weight one tall and lean and the other short and stout the former will have a greater heat production. It is possible that this explains, in part at least, why a man of thin build often eats more than a stouter man of about the same weight. Since normal adults do not differ very greatly in size and shape, it may be stated as an approximation that the heat production of the human body is one Calorie per kilogram per hour. But for the reasons just given it is much more accurate to express the basal metabolism in terms of body surface. Thus, the average basal rate of normal men between the ages of 25 and 50 years is from 40

to 37 Calories per square meter of body surface per hour. This value is constant for all normal men whether they are tall, short, thin or stout, large or small. Knowing a subject's height and weight his surface area can be determined at a glance from the chart (fig. 45 12), from the nomogram shown in figure 45 13, or from table 48, or it may be calculated from the height-weight formula.* The average surface area for adults in Canada and the United States is about 1.6 square meters for women and 1.8 for men, the total basal heat production of the majority of normal adults ranges from 1400 to 1800 Calories per day.

The heat production per square meter of body surface is arrived at by dividing the value for the total heat production per hour of the subject, as determined by one or other of the methods already described, by the figure for the surface area. For example, a man 175 cm tall weighing 75 kg has a surface area of 1.91 sq meters. His total heat production is, say, 76.20 Calories per hour. His heat production per square meter per hour is therefore

$$76.20/1.91 = 39.8 \text{ Calories}$$

* The formula introduced by Meeh and modified by Du Bois and Du Bois is as follows

$A = W^{.725} \times H^{.725} \times 71.84$ (a constant),
where A = surface area in square centimeters, W = weight in kilogram, and H = height in centimeters

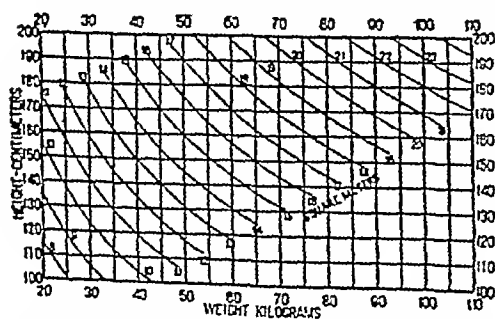


FIG 45 12 Chart for determining surface area in man in square meters (After DuBois) Example weight 60 kilograms, height 170 cm. = 1.70 sq m

Having obtained this figure it is customary to express the B M R as normal or as a percentage above (+) or below (−) the normal. Thus, in the foregoing example the rate would be said to be normal. If it were 30 Calories per square meter per hour it would be expressed as −25 per cent and if 50 Calories as +25 per cent. The age and sex must also be considered since 39.8 Calories per square meter though normal for a full-grown man of 25 years or so would be above normal for a woman of the same age and below normal for a child (see below).

CONDITIONS WHICH INFLUENCE THE BASAL METABOLIC RATE

A Physiological

(1) **AGE AND SEX.** The heat production per square meter of body surface diminishes progressively from infancy to old age being about 50 Cal per square meter per hour at the age of ten or twelve and about 32 Cal at 90 years. The metabolism of the new-born is much lower (25 Cal) than that of infants a few weeks older. Premature infants have a lower rate than those born at full term. Females have a metabolic rate a little lower than that of males in the same age group. The relationship of heat production to age and sex is given in table 49.

(2) **RACE AND CLIMATE.** Some oriental races show a slightly lower rate (from 10 to 15 per cent) than occidentals living under the same climatic conditions, others show little difference. In one oriental race (the Miao of West China) and in the Maya of Central America and the Mapuches of Chile the basal metabolic rate is actually higher than that of whites. The rate is also higher in Eskimos. The basal metabolism of white persons in a tropical climate is usually reduced.

(3) **HABITS.** Owing to the greater development of their muscular tissues athletes and laborers have in general a higher B M R than persons leading a sedentary life.

(4) **PREGNANCY.** The basal metabolic rate of the pregnant woman shows little change until the sixth or seventh month when the fetus causes an appreciable increase in the weight of the mother. The metabolism of the mother from this time to term is the sum of her own metabolism in the non-pregnant state and that of the fetus. It was found by Murlin and Carpenter, for example, that the metabolism of the new-born infant and

the post-partum metabolism of the mother added together practically equaled the metabolism of the pregnant state near the end of term. Boothby and Sandiford estimated the surface area of the fetus throughout gestation and concluded that the excess heat production of the pregnant state was derived from the fetus and the increased mass of the maternal structures, the energy production per unit of mass of the maternal organism remaining constant. Normal pregnancy, then, exerts little or no *specific* effect upon the basal metabolic rate.

(5) The nature of the **DIET** seems to have little influence upon the B M R, though in strict vegetarians it is said to be some 11 per cent lower than that of meat eaters.

(6) **VARIATIONS IN BAROMETRIC PRESSURES.** A moderate reduction in oxygen pressure does not affect the metabolic rate, but a reduction in the latter occurs, which varies by from 5 to 25 per

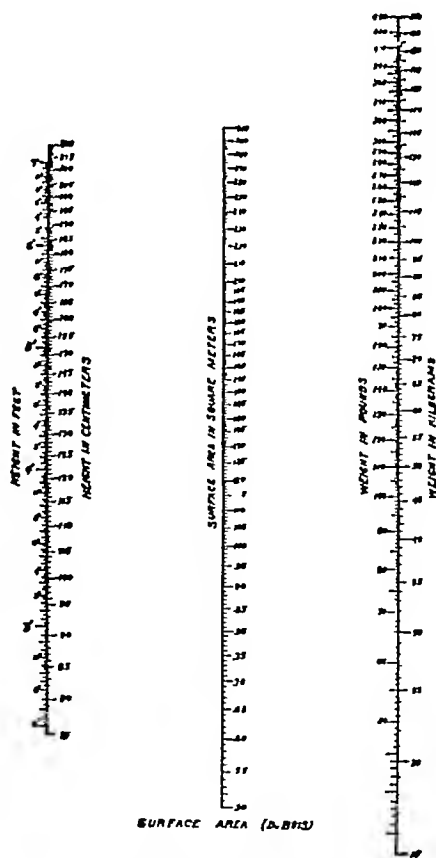


FIG 45.13 Nomogram for surface area from weight and standing height, according to the formula of Du Bois. A line joining the weight and height of an individual crosses the scale of surface area at the required value. (After Boothby and Sandiford, from Du Bois, *Basal Metabolism*, Lea and Febiger Philadelphia.)

TABLE 49

Basal and total metabolic rates of males and females
From Cruckshank compiled from Boothby and associates

AGE	B.M.R. CALORIES PER SQ METER PER HOUR		TOTAL BASAL METABOLISM IN CALORIES PER 24 HOURS		TOTAL CALORIES PER DAY	
	Males	Females	Males	Females	Males	Females
Birth	30	30	288	288	440	440
1	55	52	660	624	1000	1000
2	57	53	780	725	1200	1200
3	55	52	845	798	—	—
5	53	52	915	886	1600	1600
8	51 8	47 0	1143	993	2000	2000
11	47 2	45 2	1268	1193	2500	2500
12	46 8	43 3	1347	1267	—	—
13	46 5	42 0	1428	1330	—	—
14	46 4	41 5	1537	1391	3200	2800
15	46 0	40 0	1667	1420	—	—
16	45 7	38 8	1764	1434	3500	2600
18	43 2	37 5	1783	1440	3800	2500
20	41 6	36 3	1756	1437	3500	2400
25	40 3	36 0	1760	1442	3000	2400
40	38 0	35 0	1641	1344	3000	2400

TABLE 50

Extra Calories of metabolism per hour attributable to occupation
(After Harrop)

	EXTRA CALORIES PER HOUR
<i>Occupations of men</i>	
Tailor	44
Bookbinder	81
Shoemaker	90
Metal worker, filing and hammering	141
Painter of furniture	145
Carpenter making a table	164
Stonemason chiselling a tombstone	300
Man sawing wood	378
<i>Occupations of women</i>	
Seamstress, needlework	6
Typist, 50 words per minute	24
Seamstress, using sewing machine	57
Bookbinder	63
Housemaid (moderate work)	81
Laundress (moderate work)	124
Housemaid (hard work)	157
Laundress (hard work)	214

cent in different subjects, when the barometric pressure falls to half an atmosphere (O₂ pressure 75 mm Hg) This is around the barometric pressure at which mountain sickness usually appears (p 421) Breathing excessively high tensions of oxygen does not raise the B M R. above the normal level That is, the oxygen consumption cannot be raised simply by increasing the oxygen supply

(7) CHEMICAL SUBSTANCES, caffeine, adrenaline, thyroid extract or thyroxine, benzedrine, and dinitrophenol, raise the basal metabolic rate Smoking a cigarette, especially if the smoke is inhaled, increases the metabolism of most subjects, the average increase being around 9 per cent

B Pathological

(1) *The B M R is below normal in the following conditions*

- Starvation and undernutrition (ch 51)
- Obesity due to pituitary or hypothalamic disorders (pp 710 and 804)
- Hypothyroidism (myxedema, p 809)
- Addison's disease (p 845)
- Lipoid nephrosis (p 469)
- (2) *Conditions with a high B M R*
- Hyperthyroidism (exophthalmic goiter, p 810)
- Fever (p 739) The basal rate is raised about 7 per cent for each degree (Fahrenheit) rise in temperature
- Diabetes insipidus (p 802)
- Cardio-renal disease with dyspnea (+25 to +50 per cent)
- Leukemia (+21 to +80 per cent)
- Polycythemia (+10 to +40 per cent)

INFLUENCES WHICH RAISE THE METABOLISM OF A NORMAL INDIVIDUAL ABOVE THE BASAL LEVEL

We have already seen that the basal metabolism is defined as the heat production of a subject as nearly as possible at complete physical and mental rest, some hours after food and with the room temperature at about 20°C These three factors—muscular exercise, the ingestion of food and environmental temperature—have a powerfully stimulating influence upon metabolism

Muscular and mental effort Even light muscular exertion, e g, sitting, standing, dressing and undressing, sewing, etc., raises the metabolism by from 25 to 60 per cent above the basal level Moderate exercise, such as walking, swimming, housework, light carpentry, etc , raises it from 100

to 200 per cent, and very hard work may increase it from 10 to 15 times over the basal level. Henderson and Haggard found that in three members of the Yale University boat crew, the total heat production during rest amounted to 1 65, 1 42 and 2 4 Calories per minute respectively. The corresponding values during strenuous exercise on a rowing machine were 18 90, 21 83 and 29 37 Calories per minute. The extra energy expenditure caused by various occupations is shown in table 50. Mental effort, on the other hand, causes an almost negligible increase in metabolism. Benedict, for instance, after a series of experiments in which intense mental effort was expended in solving mathematical problems found an increase of only 3 or 4 per cent. He states, "The cloistered scholar at his books may be surprised to learn that the extra Calories needed for one hour of intense mental effort would be completely

met by the eating of one oyster cracker or one-half of a salted peanut." The *basal* oxygen consumption of the brain however, is high (p. 930), amounting to about 10 per cent of that of the entire body. *Strong emotion* may raise the metabolism from 5 to 10 per cent above the basal level. During *sleep*, when the muscles are more completely relaxed than is possible during the waking day the metabolism falls below the basal level. The reduction amounts to from 10 to 13 per cent. (Strictly speaking, the metabolism during sleep, since it is the physiological minimum, *is* the basal level.) If sleep is disturbed and associated with muscular activity the metabolism may be as high as or even higher than the basal level.

The influence of *food* and of *environmental temperature* (pp. 640 and 735) upon metabolism are dealt with elsewhere.

CHAPTER 46

PROTEIN METABOLISM

GENERAL DESCRIPTION AND CLASSIFICATION OF PROTEINS

Protein is a basic constituent of protoplasm and consequently forms a proportion of all living tissues—animals or vegetable, of some tissues, e.g., muscle, it is the predominant solid constituent. It differs from the other foodstuffs—carbohydrates and fats—in containing (in addition to carbon, hydrogen and oxygen) nitrogen, sulphur and usually phosphorus. The molecules of certain proteins contain a prosthetic group, e.g., hemoglobin, glycoproteins, lecithoproteins, nucleoproteins, etc.

A classification and brief description of the various types of protein are given in table 51

THE PROTEIN MOLECULE

The protein molecule is constructed of a number of units linked together. These units or "building stones" are called *amino-acids*. They are separated when the protein molecule is hydrolysed by boiling with acid, or by the action of the digestive enzymes (proteases). Some 25 different amino-acids have been definitely identified as constituents of the protein molecule (see table 52). Some proteins contain nearly all of these in varying proportions, in others such as gelatin there are only 14 or 15 different kinds, some *essential* amino-acids being missing, while such simple proteins as the protamines, sturine and salmine, contain only 3 and 4, respectively. The single amino-acids have molecular weights ranging from 75 for glycine to over 200 for tyrosine and nearly 800 for thyroxine. The molecular weights of those proteins composed of large aggregations of amino-acids are correspondingly great and extend over a very wide range. The molecular weight of egg albumin, for example, is 45,000, that of hemocyanin, over 6,000,000, and of tobacco mosaic virus, 40,000,000. Some basic proteins, protamines and histones, on the other hand, have molecular weights of about 2000 only.

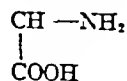
Protein molecules also vary in shape. Studies of protein structure by means of X-ray diffraction photographs reveal that the molecules of

some proteins, such as keratin, collagen and myosin—the so-called *fibrous proteins*—have an elongated fiber-like form, resulting, it is believed, from the extended arrangement of the polypeptide linkages of which the molecule is constructed (see below). When stretched such molecules become elongated still further but return to their previous length when released from the stretching force. The changes in length are described as being due to folding and unfolding of the polypeptide chains in a concertina like fashion. Other proteins—the *globular proteins*—such as egg albumin and the serum proteins, are composed of molecules which are folded or arranged into a lattice pattern to form a compact structure of a more or less globular shape. Unfolding and the assumption of a permanent extended form is associated with denaturation. Proteins of approximately the same molecular weight may differ considerably in their molecular structure and for this reason may show differences in such physical properties as elasticity, osmotic pressure or viscosity. The fundamental structure of all proteins is believed by Astbury to be fibrous, but the intermolecular arrangement of the polypeptide chains varies widely between different types and in the same protein under different conditions. Thus a variety of patterns is produced.

The average elementary composition of the molecule of a protein such as albumin or globulin is as follows: C, 54 per cent, H, 7 per cent, N, 16 per cent, S, 1 per cent, O, 22 per cent.

THE AMINO-ACIDS

The amino-acids may be regarded as derivatives of the saturated fatty acid series in which the *alpha* hydrogen atom has been replaced by an amino (NH₂) group, except in the case of proline and hydroxyproline, when the substitution is by an imino (NH) group. The simplest amino-acid is amino-acetic (glycine or glycolic)



and the structure of the other amino-acids, with

TABLE 51

CLASS OF PROTEIN	CHARACTERISTICS	EXAMPLES
<i>A Simple proteins</i>		
(1) <i>Albumins</i>	Soluble in water and coagulable by heat. Present in both animal and plant tissues	<i>Serum albumin</i> , <i>egg albumin</i> , <i>lactalbumin</i> and various vegetable albumins such as <i>leucosin</i> (in wheat, rye and barley), <i>legumelin</i> in lentils, soy-bean, beans and peas and <i>phaselin</i> in kidney bean
(2) <i>Globulins</i>	Soluble in dilute saline solutions, insoluble in water. Animal globulins are coagulated by heat. Vegetable globulins imperfectly or not coagulated by heat	<i>Serum globulin</i> , <i>fibrinogen</i> (and <i>fibrin</i>) <i>vitellin</i> of egg yolk and vegetable globulins such as <i>excelsin</i> (Brazil-nut), <i>edestin</i> (hemp), <i>phaseolin</i> (kidney bean), <i>legumin</i> (peas and lentils) and <i>tuberin</i> (potato). A number of other vegetable globulins have been isolated and named
(3) <i>Glutehns</i>	Found only in plants. Insoluble in water, saline or alcohol, but soluble in very dilute alkali	<i>Glutenin</i> of wheat, <i>oryzenin</i> of rice and <i>glutelin</i> of maize
(4) <i>Prolamines or Gliadins</i>	Found in cereals (except rice) soluble in 70-90 per cent alcohol. Insoluble in water. They contain a large proportion of proline and ammonia nitrogen	<i>Gliadin</i> of wheat, <i>hordein</i> of barley and <i>zein</i> of maize
(5) <i>Albuminoids or scleroproteins</i>	Especially resistant to the usual reagents. They enter into the construction of protective and connective tissues, e.g., skin, tendons, ligaments and bones	<i>Keratin</i> of hair, skin, bone, feathers, tortoise shell and egg-shell, <i>elastin</i> , <i>collagen</i> , <i>ossein</i> and <i>gelatin</i> of tendons, ligaments, bone, etc.
(6) <i>Histones</i>	Soluble in water and precipitated by ammonia solution and by alkaloids. They contain a large percentage of diamino acids (p. 625)	<i>Globin</i> of hemoglobin, <i>thymus histone</i> , <i>scombron</i> and <i>gadus histone</i> in spermatozoa of mackerel and cod-fish respectively
(7) <i>Protamines</i>	Found in combination with nucleic acid in heads of fish spermatozoa. Constructed predominantly of diamino acids	<i>Salmine</i> and <i>sturine</i> in spermatozoa of salmon and sturgeon respectively
<i>B Conjugated proteins</i> Proteins whose molecule is combined with another non protein group		
(1) <i>Nucleoproteins</i>	Nucleic acid in combination with a protein belonging usually to the class of histones or protamines. Found in cell nuclei	See ch. 48
(2) <i>Chromoproteins</i>	Protein in combination with a pigment (e.g., hematin) containing iron, copper or other metal	Hemoglobin, hemocyanin, etc.
(3) <i>Glycoproteins</i>	Proteins other than nucleoproteins in combination with a carbohydrate group	<i>Mucin</i> in salivary gastric and intestinal secretions, <i>ovomucoid</i> of egg white and <i>chondromucoid</i> of cartilage
(4) <i>Lipoproteins</i>	Proteins in combination with lipid	Present in plasma, milk, cell nuclei
(5) <i>Phosphoproteins</i>	Proteins other than nucleoproteins and lecithoproteins in combination with a phosphorus-containing group	<i>Caseinogen</i> (and casein), <i>vitellin</i> of egg-yolk
<i>C Derivatives of proteins—derived proteins</i> These are produced by the action of acids, alkalis or proteolytic enzymes upon certain of the proteins listed above		
(a) <i>Primary derivatives</i>		
(1) <i>Proteans</i>	Insoluble products formed in the early stage of the action upon proteins of water, dilute acids and enzymes	

TABLE 51—Continued

CLASS OF PROTEIN	CHARACTERISTICS	EXAMPLES
<i>C Derivatives of proteins—derived proteins—Continued</i>		
(2) <i>Metaproteins</i>	Formed in a later stage of the action of acid or alkali	<i>Acid metaprotein, alkali metaprotein</i>
(3) <i>Coagulated proteins</i>	Formed by the action of heat or of alcohol upon solutions of proteins	
(b) <i>Secondary derivatives</i>		
(1) <i>Proteoses</i>	Formed by the action of pepsin or trypsin upon proteins. They are soluble in water from which they are precipitated by saturation with ammonium sulphate. They are incoagulable by heat	<i>Albumose</i> from albumen, <i>globulose</i> from globulin, <i>caseose</i> from casein
(2) <i>Peptones</i>	These represent a further stage in action of proteolytic enzymes. They are soluble in water but are not precipitated from an aqueous solution by ammonium sulphate. They are not coagulated by heat	
(3) <i>Peptides, dipeptides, tripeptides and polypeptides</i>	Products formed in the final stages of proteolytic digestion	Glycyl-alanine, leucyl glutamic acid, etc.

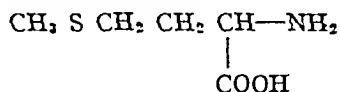
TABLE 52

Classification of the amino-acids

I. ALIPHATIC AMINO-ACIDS	
A. <i>Monoamino-monocarboxylic acids</i>	
(1) Glycine (or glycolic acid) C_2H_5NO , or amino-acetic acid	(6) Norleucine, $C_6H_{13}NO$, or α amino- <i>n</i> -caproic acid
$\begin{array}{c} CH-NH_2 \\ \\ COOH \end{array}$	$\begin{array}{c} CH_2, CH_2, CH_2, CH-NH \\ \\ COOH \end{array}$
(2) Alanine, C_3H_7NO , or α amino-propionic acid	(7) Leucine, $C_6H_{13}NO$, or α amino-isocaproic acid
$\begin{array}{c} CH_2, CH-NH_2 \\ \\ COOH \end{array}$	$\begin{array}{c} CH_3 \\ \diagdown \\ CH \\ \diagup \\ CH_3 \end{array} \begin{array}{c} CH_2, CH-NH_2 \\ \\ COOH \end{array}$
(3) Serine, C_3H_7NO , or α -amino- β -hydroxy propionic acid	(8) Isoleucine, $C_6H_{13}NO$, or α amino- β -ethyl- β -methyl propionic acid
$\begin{array}{c} CH \quad CH-NH_2 \\ \quad \\ OH \quad COOH \end{array}$	$\begin{array}{c} CH_3 \\ \diagdown \\ CH \\ \diagup \\ CH_2, CH- \end{array} \begin{array}{c} CH-NH_2 \\ \\ COOH \end{array}$
(4) Threonine, C_4H_9NO , or α amino- β -hydroxy- <i>n</i> -butyric acid	<i>Sulphur-containing monoamino-monocarboxylic acids</i>
$\begin{array}{c} CH_2CH \quad CH-NH_2 \\ \quad \\ OH \quad COOH \end{array}$	(9) Cystine, $C_6H_{11}N_2S_2O_4$, (or di-cystine) or di-(β -thio- α amino-propionic) acid
(5) Valine, $C_5H_{11}NO$, or α amino-isovaleric acid	$\begin{array}{c} CH_2-S-S-CH_2 \\ \quad \\ CH-NH_2 \quad CH-NH_2 \\ \quad \\ COOH \quad COOH \end{array}$
$\begin{array}{c} CH_3 \\ \diagdown \\ CH \\ \diagup \\ CH_3 \end{array} \begin{array}{c} CH-NH \\ \\ COOH \end{array}$	

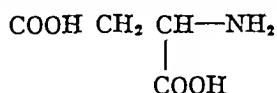
TABLE 52—Continued

- (10) Methionine,
- $C_5H_{11}SNO_2$
- , or
- α
- amino-
- γ
- methylthiol-
- n*
- butyric acid

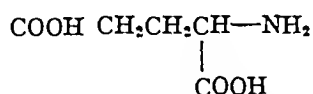


B Monoamino-dicarboxylic acids

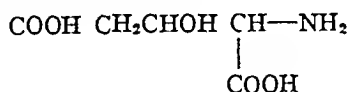
- (11) Aspartic acid,
- $C_4H_7NO_4$
- , or
- α
- amino-succinic acid



- (12) Glutamic acid,
- $C_5H_9NO_4$
- , or
- α
- amino glutaric acid

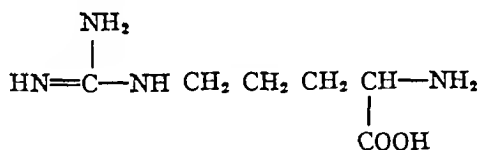


- (13) Hydroxyglutamic acid,
- $C_5H_9NO_6$
- , or
- α
- amino
- β
- hydroxy-glutaric acid

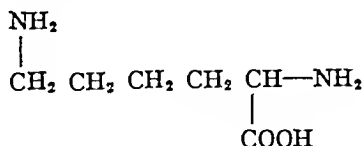


C Diamino-monocarboxylic acids

- (14) Arginine,
- $C_6H_{14}N_4O_2$
- , or
- α
- amino-
- δ
- guanidine-
- n*
- valeric acid

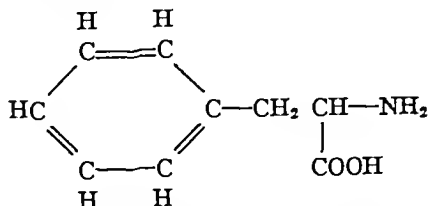


- (15) Lysine,
- $C_6H_{14}N_2O_2$
- , or
- α
-
- ϵ
- diamino-caproic acid

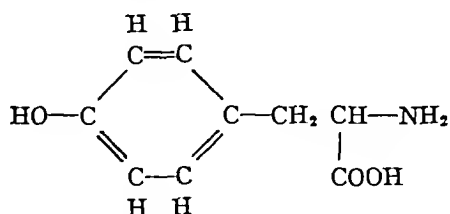


II AROMATIC AMINO-ACIDS

- (16) Phenylalanine,
- $C_9H_{11}NO_2$
- , or
- α
- amino-
- β
- phenylpropionic acid

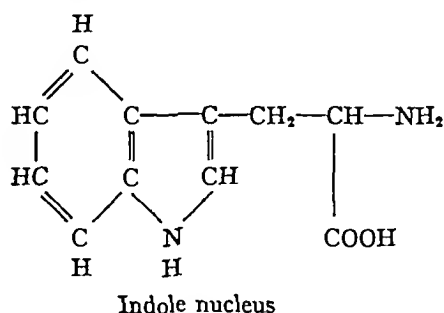


- (17) Tyrosine,
- $C_9H_{11}NO_3$
- , or
- α
- amino-
- β
- parahydroxyphenyl-propionic acid

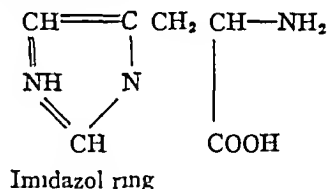


III HETEROCYCLIC AMINO ACIDS

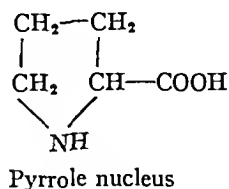
- (18) Tryptophane,
- $C_{11}H_{12}N_2O_2$
- , or
- α
- amino
- β
- indolepropionic acid



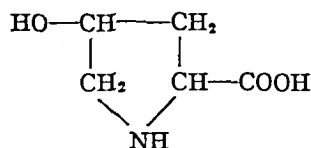
- (19) Histidine,
- $C_6H_9N_3O_2$
- , or
- α
- amino-
- β
- imidazolpropionic acid



- (20) Proline,
- $C_5H_9NO_2$
- , or pyrrolidine-
- α
- carboxylic acid

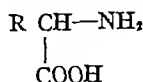


- (21) Hydroxyproline (oxyproline),
- $C_5H_9NO_3$
- , or
- γ
- hydroxy-pyrrolidine-
- α
- carboxylic acid

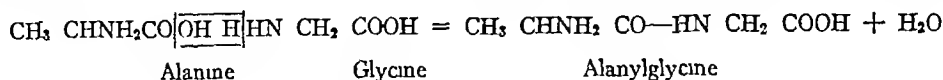


Thyroxine (p 821), 3-5 diiodotyrosine or iodogorgoic acid, citrulline (α -amino- δ -carbamido-*n*-valeric acid) and ornithine (α , δ -diamino valeric acid), and several others, would have to be added to complete the list of naturally occurring amino-acids

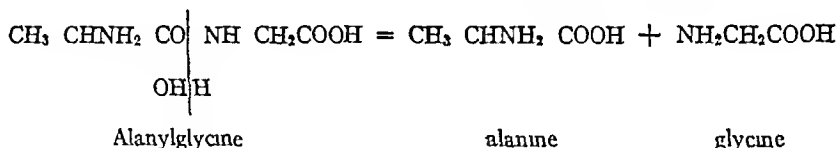
the exception of proline and oxyproline, may be represented by the following type formula



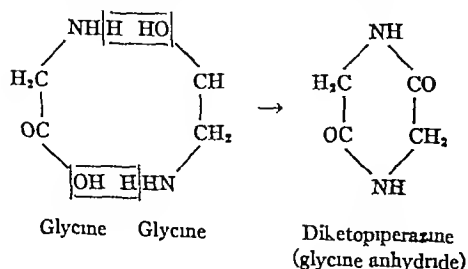
Amino-acids therefore contain a basic (NH_2) group and an acid (COOH) group R represents a chemical group which varies greatly in size and structure. In the synthesis of the protein molecule chains of amino-acids are formed by the linkage of the basic group of one amino acid with the acid group of another and the liberation of a molecule of water. Thus



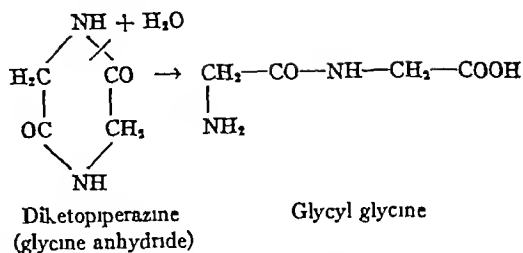
The junction, $\text{CO} \text{NH}$, whereby amino-acids become grouped together is called the *peptide linkage*. The reverse process, namely, the separation of amino-acids from one another is also effected at this link in the chain, a molecule of water first being taken up. This process, which is called *hydrolysis*, may be illustrated thus



Besides the chain like combination of the amino-acids in the protein molecule, a smaller proportion are believed to be united to form closed ring compounds,—*diketopiperazines* (*amino-acid anhydrides*). The amino-acids undergo dehydration and unite by their free amino and carboxyl groups to form the diketopiperazine ring. In the example shown below two glycine molecules join to produce glycine anhydride.



Upon hydrolysis by heating with acid the ring is broken, and in the case of glycine anhydride the dipeptide glycyl glycine is formed



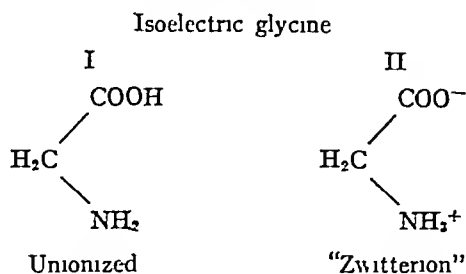
Protein, as a consequence of its constitution of amino-acids linked together as described above, contains free NH_2 and COOH groups. By virtue of these it can act either as a weak base or as a weak acid. In acid solution it acts as a base yielding cations to form protein chloride, sulphate, etc. When a current passes

through the solution the amino-acid cation migrates to the cathode or positive pole.¹ In alkaline solution it behaves as an acid, forming anions to produce a proteinate of sodium, potassium etc.,. In an electric field the amino acid anion migrates to the anode. On account of these opposite reactions, depending upon the acidity or alkalinity of the solution, amino-acids and

the proteins which they compose are called *amphoteric electrolytes* or *ampholytes* (Gk. *amphō*, both). At a certain hydrogen ion concentration, which varies rather widely between different proteins, the amino-acids are electrically neutral and protein behaves neither as an acid nor as a base. In an electric field no migration either to anode or cathode occurs. The pH at which this occurs is known as the "*isoelectric point*" though the effect rather than being restricted to a point extends over a pH range which with some proteins is considerable. *Isoelectric zone* is therefore a better term (table 53). Within the isoelectric zone the solubility of the protein is least and is readily precipitated by alcohol, neutral salts and other reagents. Some proteins are quite insoluble at the isoelectric point and precipitate spontaneously. According to the classical conception the electrical neutrality of protein at the isoelectric point was due to ionization of the amino-acids being at a minimum or absent, as represented below in formula I. This view has given

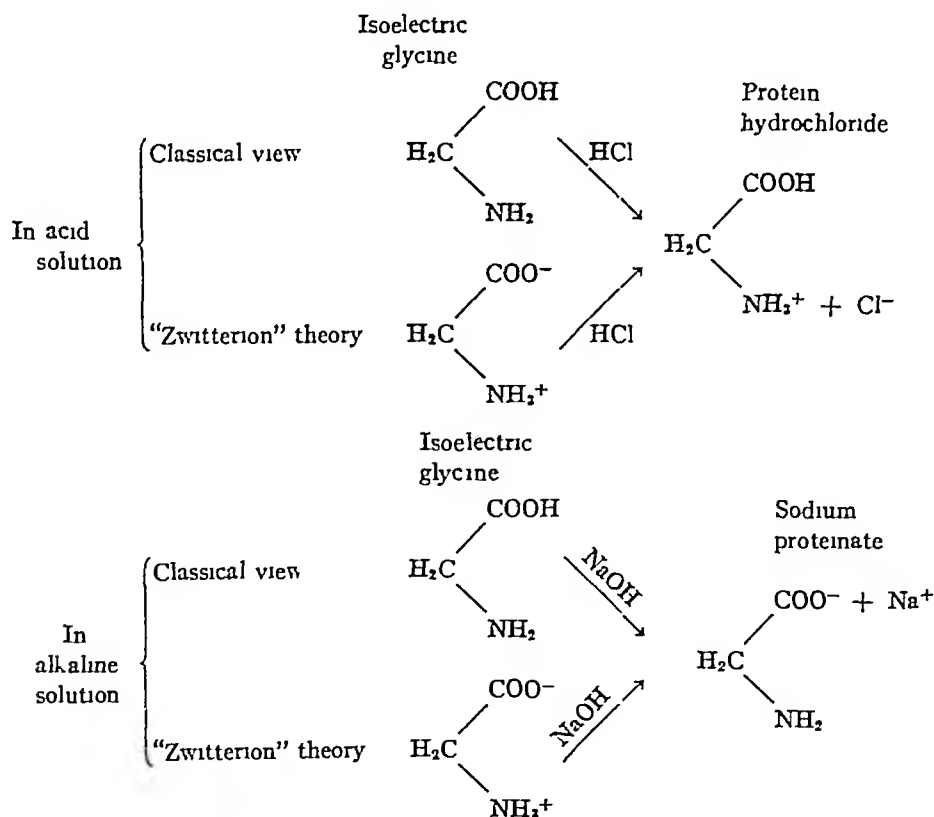
¹ This migration in an electric field of cation and anions to the respective poles is called *electrophoresis* or *iontophoresis*.

place to the *Zwitter-ion theory*, which more readily explains the experimental findings



The "*Zwitterion*" theory This postulates that the amino-acids at the iso-electric point carry equal numbers of positive and negative charges and this accounts for their electrical neutrality They are called "*zwitterions*" (Ger *Zwitter*, hermaphrodite) or *dipolar ions* (see formula II above) Much experimental support has been gained for this theory since it was first proposed

The behavior of an amino acid (glycine) in acid and in alkaline solution according to the two conceptions is illustrated below



It will be seen that the end result on either side of the isoelectric point is the same according to either view According, however, to the *Zwitter hypothesis* the ampholyte molecule gives off, at the isoelectric point, *equal numbers* of basic and acid ions, thus leaving ions—the so-called *zwitterions*—holding equal numbers

TABLE 53

Approximate isoelectric points of some amino acids

AMINO-ACID	APPROXIMATE ISOELECTRIC POINTS, pH
Glycine	6.1
Alanine	6.1
Valine	6.0
Leucine	6.0
Glutamic acid	3.2
Aspartic acid	3.0
Arginine	10.5
Lysine	9.6
Histidine	7.4
Phenylalanine	5.4
Tyrosine	5.4

of negative and positive charges (formula II) Many proteins, e.g., metaproteins, casein, etc., are almost insoluble at the isoelectric point—*isoelectric precipitation*

THE ABSORPTION OF PROTEIN

Under ordinary circumstances only negligible amounts of unchanged protein, or even such of its derivatives as proteoses, peptones and polypeptides, are absorbed into the blood from the ali-

mentary tract. Generally speaking the protein molecule must first be hydrolyzed into its constituent amino acids. These are absorbed from the small intestine but not from the stomach to any significant extent in so far as nutrition is concerned. They enter the circulation mainly through the portal blood, much smaller amounts pass into the lymph (chyle) and hence into the thoracic duct.

Small quantities of certain proteins, e.g., raw egg white and blood serum are sometimes absorbed from the intestine into the blood stream. They are excreted in the urine to a large extent unchanged. Such absorption which occurs more readily in young animals and children, should be regarded as a defect rather than as a physiological process. Nevertheless, experiments have shown that small quantities of protein introduced parenterally may be utilized by the tissues, and may even be capable of maintaining nitrogen equilibrium (p 638). Though it is possible that the injected protein was broken down into its constituent amino-acids by the general tissue cells, it is more probable that this function was performed by the cells of the intestinal wall, and that the latter are necessary for the digestion even of protein introduced parenterally. It is, however, only for a short period and to a very limited extent that the body can utilize injected protein, for proteins are species specific and after a number of injections immunity is established to the foreign protein, an antibody known as a *precipitin* is formed which causes its precipitation. Furthermore, if a subsequent injection is given two weeks or so after a single injection or after the last of a series of such injections, a most serious toxic state—*anaphylaxis*—may ensue and prove fatal (p 306). In the human subject such anaphylactic (or anaphylactoid) reactions may follow the second injection of a foreign protein, e.g., diphtheria antitoxin (horse serum) or antitetanic serum. Skin rashes such as urticaria, erythema, eczema, giant edema, etc., or pain and swelling of the joints may occur. Collapse and death occasionally result. The sensitivity of some persons to certain proteins is also sometimes responsible for dietary idiosyncrasies. Certain foods such as shell-fish, milk, strawberries, celery, etc., when ingested by persons sensitive to the proteins contained in these substances, result in cutaneous eruptions and sometimes localized edema. Minute amounts of the protein apparently enter the blood stream in an unchanged or partially digested state. Asthma, hay fever and other allergic conditions (see also p 428) have been traced to foreign proteins (pollen, cat or horse hair, feathers, etc.) entering the body through the respiratory passages.

Homologous plasma protein, on the contrary, does not cause these effects and Whipple and his associates found that dogs could be maintained

in nitrogen equilibrium (p 638) by the intravenous injection of plasma from other dogs as the sole source of protein.

A suitable assortment of amino-acids, such as an hydrolysate of casein supplemented by cystine and tryptophane, injected intravenously, is capable of maintaining an animal in nitrogenous equilibrium. Animal gelatin or isinglass (fish gelatin) which differs from complete proteins in being non antigenic, might be used for this purpose if supplemented by the missing amino-acids.

THE FATE OF AMINO-ACIDS AFTER ABSORPTION

DEAMINATION This term means the removal of amino groups from the amino-acids. Less than 20 per cent of the amino acids absorbed into the portal blood pass through the liver unchanged into the general circulation. A proportion of these are excreted intact in the urine, some are deaminated by the kidney and the nitrogen excreted as ammonia, others are utilized without alteration by the tissues. The greater part of the amino-acids reaching the liver are there retained and deaminated. The ammonia which is split off combines with carbon dioxide to form urea as described on page 633. The deaminated remainder, i.e., the fatty acid residue of the amino acid molecule may either undergo oxidation, and thus furnish energy to the body, or be transformed into glucose. The glucose may be burned or be stored as glycogen, or again it may be transformed into fat. Not all amino-acids, however, are sugar or glycogen formers. Those which play the chief role in this respect are *glycine*, *alanine*, *aspartic*, *glutamic* and *hydroxyglutamic acids*, *serine*, *cystine*, *hydroxyproline* and *proline*. With the exception of *hydroxyproline* and *proline* these are all straight chain amino acids with less than six carbon atoms. *Arginine*, *valine*, and *threonine* form sugar to a limited extent, whereas, *leucine*, *isoleucine*, *lysine*, *methionine*, *tryptophane*, *histidine*, *tyrosine* and *phenylalanine* are not glycogenic (see table 54). The quantity of amino-acids present in 100 grams of protein is sufficient to form some 58 grams of glucose.² A depancreatized or phlorid-

² Since different proteins contain varying amounts of sugar-forming amino-acids, this is an average figure. It was arrived at by calculation from the proportion of glucose (dextrose) to nitrogen in the urine—the so-called G N (or D N) ratio—in phloridized dogs fed exclusively upon protein or during a period of starvation. Lusk found the G N ratio under such circumstances to be 3.65:1. The nitrogen of the urine is derived, of course, practically entirely from protein.

zimized dog, for instance, or the subject of severe diabetes, though upon a carbohydrate-free diet, continues to excrete large quantities of glucose. Also, the normal animal during prolonged starvation maintains its blood sugar at practically the normal level, which suggests that body protein is undergoing conversion to glucose. When protein is fed after a period of starvation, glycogen accumulates in the liver. There is also evidence that the conversion of protein to glucose occurs even in a normal animal upon a mixed diet. The conversion process probably occurs only in the liver. Mann and Magath found that the hypoglycemia following hepatectomy was not affected by the intravenous injection of glycine. The conversion process is influenced by the anterior lobe of the hypophysis and the adrenal cortex.

Though deamination is a function of the kidney and other tissues as well, its main site is the liver. In the dog it appears to be carried on exclusively by hepatic tissue. Mann and his colleagues found no evidence in dogs of extrahepatic deamination. The injection of amino-acids into the blood stream of a normal dog caused a rapid rise in the amino-acid concentration of the blood and increased amino-acid excretion in the urine. The level in the blood returned to normal in about two hours. *About 25 per cent of the injected amino-acid nitrogen appeared in the urine as urea.* After hepatectomy, the urea concentration of the solid tissues, blood and urine fell, while the concentration of amino-acids rose, amino-acids injected into the blood stream were *completely recovered* unchanged, about a third from the urine and the remainder

The urinary glucose of an animal under the influence of phloridzin and whose glycogen stores have been exhausted is also assumed to be derived exclusively from protein. Now each gram of urinary nitrogen represents catabolism of 6.25 grams of protein. Therefore with a G/N ratio of 3.65:1 every 3.65 grams glucose secreted indicates the conversion of 6.25 grams of protein. So

$$(3.65/6.25) \times 100 = 58 \text{ per cent}$$

In diabetic (depancreatized) animals during starvation or upon a protein diet the G/N ratios obtained by different observers vary widely. Minkowski's figure of 2.8:1 is most usually quoted (this would indicate the conversion of 45 per cent of protein to sugar). Macleod and his associates, however, have obtained ratios in depancreatized dogs after the withdrawal of insulin as high as 6:1 in some experiments, and less than 2:1 in others. The ratio did not show a constant value, either in different animals or in the same animal at different times. They therefore seriously question the reliability of the G/N ratio as an index of the extent of the protein to glucose conversion, and consider that protein is not the only source of the urinary glucose.

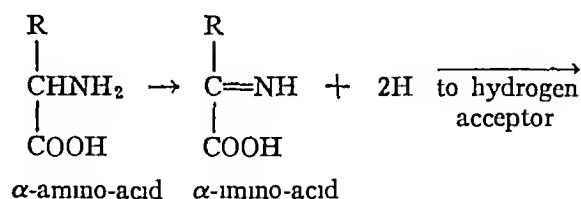
TABLE 54

Classification of amino-acids according to their glyco-genic or ketogenic properties

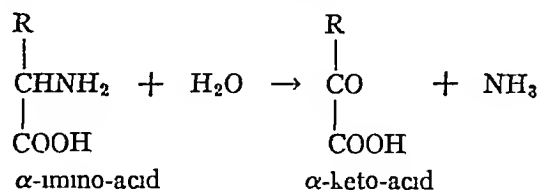
GLYCOGENIC	KETOGENIC	NEITHER GLYCOGENIC NOR KETOGENIC
Glycine Alanine Serine Threonine Cysteine Aspartic acid Valine Glutamic acid Arginine Ornithine Proline Hydroxyproline	Leucine iso-Leucine Phenylalanine Tyrosine	Norleucine Methionine Lysine Tryptophane Histidine

from the muscles. There was no sign of deamination or of urea synthesis.

Deamination is generally believed to be an oxidative process brought about by various enzymes in liver, kidney, and to some extent in other tissues. Oxidation of the α -carbon atom, i.e., the carbon atom to which the NH_2 group is attached, first occurs with the production of an α -imino acid. This involves the transference of two hydrogen atoms to a hydrogen acceptor through the action of *D-amino-oxidase*, *glycine oxidase*, or of *L-glutamic dehydrogenase*.



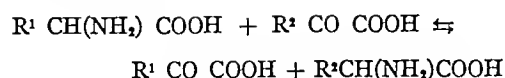
The imino-acid then undergoes spontaneous hydrolysis to form an α -keto-acid



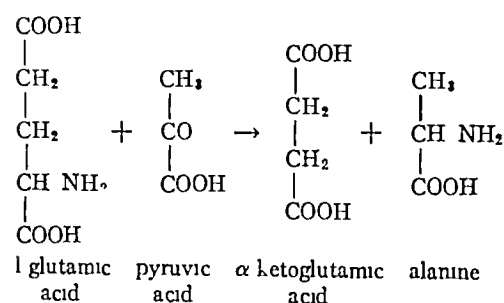
TRANSAMINATION An important reaction occurring in the metabolism of protein is the transference of the amino group from an α -amino acid to an α -keto acid. The change is catalyzed by an enzyme *transaminase*.

The reaction is reversible and occurs according

to the equation



Transamination was first suggested as a possibility by Needham and later demonstrated in pigeon breast muscle by Braunstein and Kritzmann. They found that when l glutamic acid and pyruvic acid were incubated with chopped liver or muscle α ketoglutaric acid and alanine were produced, the amino group of the glutamic acid being transferred to pyruvic acid



According to these investigators the majority of the amino acids (lysine and ornithine are exceptions) can take part in transamination reactions, their NH_2 groups being transferred to other amino-acids or to various keto-acids. Glutamic acid is especially active in this respect. Transamination occurs not only in liver and muscle but has been demonstrated in a number of other tissues and appears to be a function of tissues in general. Krebs observed that glycolysis in brain and retina was inhibited by glutamic acid, an effect which has since been shown to be due to the conversion of pyruvic acid to alanine through the transference of amino groups from the glutamic acid. Since transamination is a rapid process and concerned chiefly with those compounds, which play key rôles in intermediary metabolism, Cohen suggests that it represents a "shuttle" mechanism in tissue respiration "whereby certain key protein and carbohydrate intermediates are rapidly interconverted." The demonstration of transamination affords further evidence for the highly dynamic character of protein metabolism. Amino-groups are being continually exchanged between nitrogenous and non-nitrogenous compounds, and amino-acids thereby synthesized. When an amino-acid into which isotopic nitrogen (N^{15}) has been incorporated is fed and the tissues of the animal analyzed a large

number of amino acids are found to contain the isotope. Even ammonium citrate containing N^{15} fed to rats upon a low protein diet, is utilized for amino-acid synthesis. Aspartic acid, histidine, arginine, glutamic acid and proline, isolated from the animals' bodies, were found to contain the isotope.

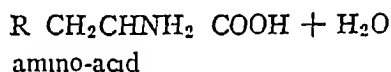
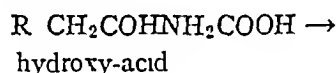
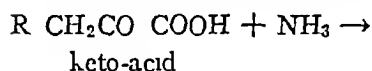
TRANSMETHYLATION Though it has long been known that certain compounds when fed are excreted by the kidney in methylated form, e.g., pyridine as N-methyl pyridine, the mechanism involved in the transference of the methyl group, from one compound to another, has only been discovered comparatively recently by du Vigneaud and his associates. Homocysteine can be converted in the body to methionine by the addition of a methyl group. The methyl group is furnished by choline or betaine. An animal fed upon a diet lacking in methionine, but containing homocysteine fails to grow, growth is resumed upon the addition of choline or betaine, the essential amino-acid methionine being thus formed. The transference of the methyl groups from methionine or choline to glycocyamine with the formation of creatine is described on page 636. Methionine, choline and betaine are, therefore, spoken of as transmethylation agents. Another example of transmethylation is the excretion of nicotinic acid, in part as trigonelline (p 752). The methyl group is not improbably derived from methionine.

Transamidation This is the transference of an NH_2 amidine group $\begin{array}{c} \text{NH}_2 \\ | \\ \text{C} = \text{NH} \end{array}$ as from arginine to glycine to form glycosamine (p 636).

THE SYNTHESIS OF AMINO-ACIDS

That the body tissues, given the necessary amino-acids, can link them together and so synthesize protein is, of course, unquestioned. The extent to which the amino acids themselves can be synthesized has been a more difficult question to decide. That synthesis of amino acids does occur is clearly indicated by the fact that only a limited number (10) of amino-acids found in animal tissue are essential constituents of the diet (p 646). Undoubtedly the body is able to synthesize glycine—the simplest amino-acid. For example, milk proteins contain no more than 0.1 to 0.3 per cent of glycine, yet from 100 grams of the former the suckling animal can build up 78 grams tissue protein containing 2.5 grams of glycine (Magnus-

Levy) Also, the liver and kidney detoxicate benzoic acid by combining it with glycine to form hippuric acid which is excreted in the urine. When large quantities of benzoic acid are fed, the glycine in the excreted hippuric acid is greater in amount than that which could have been supplied preformed from body tissue. Analogous experiments with the detoxication of phenylacetic acid to phenylacetyl glutamine indicate that glutamic acid can also be synthesized. The mode by which synthesis of the amino-acids takes place is obscure. The formation of amino-acids by the transference of amino-groups to non-nitrogenous compounds, e.g. pyruvic acid formed in the breakdown of carbohydrate, or even the utilization of ammonia, fed in the form of ammonium salts, has been mentioned. The ammonia probably unites with a keto-acid to form an aminated hydroxy-acid which by reduction yields an amino-acid.



Emblen found that alanine was formed by the surviving liver perfused with blood to which the ammonium salt of pyruvic acid had been added. Phenylalanine and tyrosine were also formed when the liver was perfused with the ammonium salts of the corresponding keto-acids. Even the addition of ammonium chloride to the fluid perfusing a glycogen-rich liver resulted in the production of alanine. The latter also followed perfusion with ammonium lactate. Furthermore, it was shown by Knoop that when γ -phenyl- α -keto butyric acid was administered to dogs the corresponding amino-acid appeared in the urine, while the experiments of Cox and Rose and of Sherwin and their associates indicate that histidine, an essential amino acid, can be formed in the body when imidazole pyruvic acid or imidazole lactic acid is fed. The synthesis of arginine has also been demonstrated by Rose and his colleagues.

To sum up it is established that glycine can be synthesized from precursors already present in the body. The same may also be said for glutamic and aspartic acids, alanine and arginine. The perfusion experiments just mentioned suggest that alanine may be synthesized from the derivatives of carbohydrate metabolism plus the ammonia which would otherwise be excreted as urea. Other

more complex amino-acids, e.g., tryptophane, histidine, and proline, etc. (p. 625) contain special groupings which cannot be formed in the body. They are therefore essential constituents of the diet (p. 645). If, however, the special groups are supplied, as in the form of the keto- or hydroxy-acid, the corresponding amino-acid can in some instances be formed.

THE SYNTHESIS OF PROTEIN

The biochemical processes leading to the synthesis of protein are very poorly understood. Borsook and Wasteney succeeded in synthesizing *in vitro* a protein-like material called "plastein" by the incubation of a concentrated peptic digest with pepsin. The pepsin thus acted reversibly, that is, induced a reaction the opposite to hydrolysis, which resulted in the recombination of the protein degradation products. No doubt the conditions of this experiment bear only a slight resemblance to those existing within the cells, but it suggests that enzymes (proteinases) acting reversibly in a somewhat similar fashion may bring about the synthesis of tissue protein.

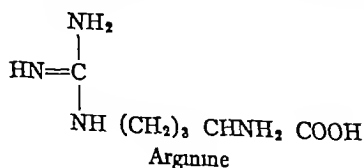
THE CLASSICAL AS OPPOSED TO THE MODERN VIEW OF PROTEIN METABOLISM

The classical theory of protein metabolism as proposed by Folin has held the field for many years. It was based upon the urinary excretion of nitrogen. Folin pointed out that under widely altered intakes of nitrogen the quantities of creatinine and neutral sulphur excreted in the urine remained remarkably constant, whereas the excretion of urea varied in amount. Creatinine and neutral sulphur were thought to be derived from the "wear and tear" of body tissue and were taken as an index of *endogenous* protein metabolism. Urea, on the other hand, was believed to be derived solely from the catabolism of dietary protein and was referred to as the *exogenous* quota of protein metabolism. Several facts, discovered chiefly through experiments with isotopic nitrogen (N^{15}) used as a "tracer", are not in accord with Folin's theory. Body structure does not remain fixed and unchanged except for a certain loss from wear and tear which is repaired from dietary material, but is constantly in a state of flux. It has been shown, for example, that when amino-acids containing isotopic nitrogen are fed to an animal in nitrogen equilibrium only small quantities of isotopic nitrogen appear in the urine,

whereas, a relatively large amount (about half) can be recovered from the tissues. In other words, the dietary nitrogen had replaced catabolized tissue protein whose nitrogen was excreted in the urine. The nitrogen derived from tissue protein, thus, far exceeded in amount that which the classical theory attributes simply to wear and tear. It has also been found, contrary to expectations, if creatinine and neutral sulphur excretion an index of endogenous protein catabolism, that increasing the metabolic rate or that of the endogenous protein catabolism does not increase significantly the urinary excretion of these substances. It was shown by Schoenheimer and his associates, using isotopic amino-acids, that dietary nitrogen was constantly and rapidly being incorporated into plasma proteins and into the proteins of liver, kidney and intestinal tract as well as, but more slowly, into the proteins, including hemoglobin, of the red cells. Nitrogen is continually being exchanged between the various tissues. Their constituents are in a state of dynamic equilibrium and constantly "turning over." In the new theory originally proposed by Borsook and Keighley, protein metabolism is conceived as a dynamic mechanism in which breakdown and resynthesis proceed hand and hand, dietary protein replacing tissue nitrogen and the nitrogen of different organs undergoing continuous interchange. Even in the starving animal tissue protein does not undergo catabolism alone, but is being continuously resynthesized, new protein being formed in one tissue from nitrogen derived from another.

Protein storage—labile protein

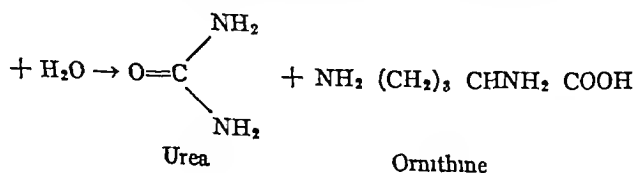
Though it is accepted beyond all question that fat and carbohydrate are stored, it was not the belief that the body accumulated reserves of protein. (The increase in body protein which occurs during growth, pregnancy or athletic training cannot be considered as protein storage in the ordinary sense.) It has been shown, however, that in a sense protein storage also occurs. During starvation or upon a non-protein diet this *reserve protein* or *labile protein* is drawn upon. When protein is fed after a period of protein deprivation,



retention of nitrogen occurs as a result of the reaccumulation of the protein stores. After these have been replenished the quantity of nitrogen excreted in the urine again balances that taken in the food. There is thus a lag in the establishment of nitrogen equilibrium in changing from one protein level to another. The liver and kidney, and especially the former, are capable of increasing considerably their content of protein. But all organs and tissues including even the blood plasma contain variable amounts of protein at different levels of protein intake. During a fast of a few days, the liver may lose 20 per cent of its protein, and when an animal in nitrogenous equilibrium is placed upon a high protein diet the balance becomes positive for a few days. The retained nitrogen is accumulated mainly in the liver. There is no reason to believe that the so-called protein reserve differs chemically or functionally from the protein characteristic of the particular organ or tissue in which it is laid down. The plasma proteins are also thought to constitute a reserve store of protein upon which the body can draw during a period of protein starvation. According to Boothby and associates, the myxedematous swelling characteristic of hypothyroidism (ch 58) is due to an increase in the quantity of protein stored in the subcutaneous tissues and skin. The excessive deposit is broken down under the influence of thyroxine. The latter also reduces the quantity of deposit protein in the normal subject.

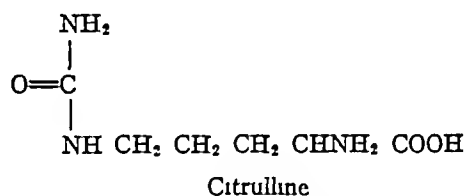
THE END PRODUCTS OF PROTEIN METABOLISM, UREA

The great proportion of the nitrogen released by the catabolism of the amino-acids appears in the urine of man, mammals, amphibia and clasmobranch fish as urea. In birds and reptiles, on the other hand, the chief end product of protein breakdown is *uric acid*. In man and other mammals the latter is derived from the metabolism of the purines (ch 48). Urea is formed directly (i.e., without preliminary deamination), from *arginine*, which is hydrolyzed by the action of an enzyme—*arginase*—into *urea* and *ornithine*. Arginase is present in the liver of mammals but not in that of birds and reptiles. The reaction may be represented thus

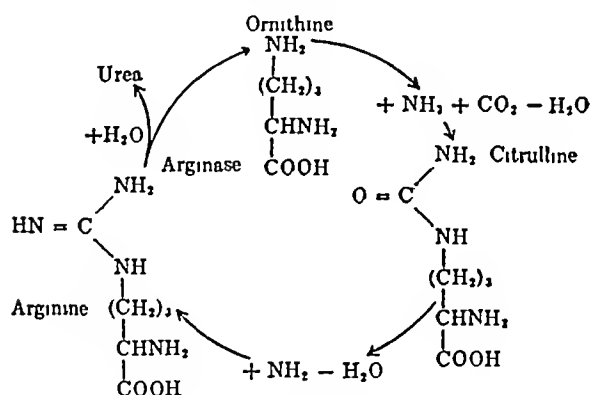


It is only within recent years that the importance of this mechanism in the production of urea has been realized and its details disclosed. It has been thought to account for the greater proportion of the urea formed in the body. When ornithine was incubated with slices of liver in the presence of ammonia and carbonic acid, large amounts of urea were formed. It is supposed that the ornithine combines with the ammonia and carbonic acid to form arginine which in turn is hydrolyzed by arginase into urea and ornithine. Thus arginine serves as an intermediary in the production of urea from the ammonia supplied by other amino-acids. The ornithine liberated by the decomposition of arginine is used over again. Ornithine thus acts after the manner of a catalyst to facilitate urea production.

According to Krebs and Henseleit, the formation of arginine from ornithine occurs in two steps. (1) The formation of the amino-acid *citrulline* by the addition of a molecule each of ammonia and carbon dioxide to ornithine.



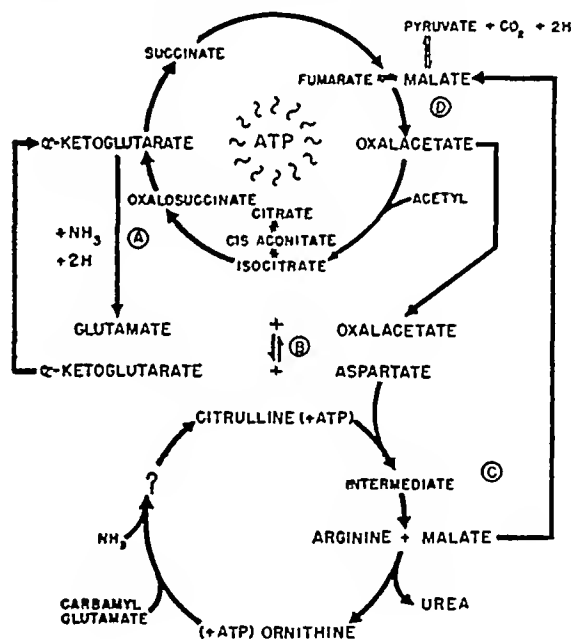
(2) The addition of a second molecule of ammonia to citrulline with the production of arginine. Therefore, the production of urea from ammonia and carbonic acid may be represented thus:



The ornithine cycle

Though the ornithine cycle is probably not the only mechanism involved in the production of urea, it is generally considered to be the major one. Some of its details have been clarified by the work of several investigators (Cohen and Hayana, Borsook and Dubnoff, and Ratner and associates). Cohen

and Hayana reported that for the production of arginine from citrulline glutamic acid rather than ammonia appeared to be the nitrogen donor, and that Mg ions, adenosinetriphosphate (ATP) and oxygen were required for the reaction. The transfer of the NH₂ group it has been suggested, is effected by transamination and the simultaneous loss of two H atoms (transamination). The citrulline-arginine conversion occurred only in the presence of oxygen. Glutamate also acts as an acceptor of CO₂ and NH₃ in the synthesis of citrulline from ornithine, in the first stage of the ornithine cycle. Aspartic acid is, however, according to the researches of Ratner and her colleagues, the specific nitrogen donor in the conversion of citrulline to arginine, which occurs anaerobically in the presence of Mg ions and ATP. Malic acid is also formed in this reaction, but undergoes oxidation to oxaloacetic acid. Through this latter reaction the tricarboxylic and ornithine cycles are interlocked. The following scheme is given by Ratner:



Urea is decomposed into NH₃ and CO₂ by the action of urease. This enzyme is found in leguminous plants, especially in soy bean, and in relatively high concentration in the gastric mucosa (ch 39). Urease is employed in the determination of urea (from ammonia evolved) in urine and in other body fluids and in tissues.

THE EXCRETION OF UREA AND ITS DISTRIBUTION IN THE BODY. On an ordinary mixed diet from 80 to 90 per cent of the urinary nitrogen is urea-nitrogen. The absolute amount of urea-nitrogen excreted daily is usually from 9 to 13 grams (20

to 30 grams of urea) Minimal amounts are also excreted in the sweat (ch 53), salivary, intestinal and mammary secretions The urea-nitrogen varies with the protein content of the diet. Upon a low protein diet the output may be as low as 2 grams and on a diet rich in protein over 25 grams The value of urea-N is therefore taken as an index of the magnitude of the catabolism of food protein (exogenous metabolism, see table, p 637)

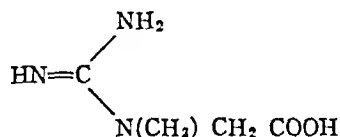
The blood contains from 8 to 15 mg of urea-N (18 to 35 mg urea) per 100 cc. Urea is readily diffusible and is found in about the same concentration in the various tissues and other fluids, e g, lymph, bile, cerebrospinal fluid and pancreatic juice, as in blood The kidney, however, is an exception since it contains some 150 to 200 mg of urea per 100 grams of tissue We have already seen that the urinary ammonia is formed from glutamine and amino acids in the kidney and not from urea (p 458)

THE SITE OF UREA PRODUCTION It has long been known that the liver is an important site of urea formation Perfusion of the liver with amino-acids or their addition to sliced liver tissue leads to the production of urea Urea is very widely distributed throughout nature, being a constituent of the body fluids of the lowest forms of animal life It is also found in plants For this reason its formation in animals might be thought to be a function common to all tissues But the experiments of Mann and his associates in hepatectomized dogs, points to the liver as being the main if not the sole site of urea production After complete removal of the liver the formation of urea ceased, so that its concentration in the blood and urine fell progressively If the kidneys as well as the liver were excised there was no accumulation of urea in the blood though the amino-acid concentration rose Also, after the injection of glycine or alanine into the hepatectomized animal, in contrast to the result obtained in the normal animal, the concentration of the injected amino acid rose in the blood and urine, and there was no evidence of urea production That these results in dogs apply to man is indicated by the observation of Rabinowitch He has reported that in a case of extensive damage to kidneys and liver (acute yellow atrophy) the excretion of urine was almost suppressed yet the blood was urea-free The appearance of leucine and tyrosine in the urine which occurs in acute yellow atrophy of the liver may also be ascribed to the failure of deamination of these amino-acids Never-

theless, though the liver undoubtedly plays the predominant rôle in the production of urea, certain other tissues may under certain circumstances share to some degree in its formation Graham and associates found, for example, that the active mammary gland produces not inconsiderable amounts

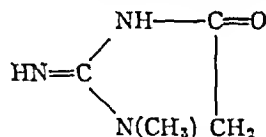
CREATINE AND CREATININE

Creatine ($C_4H_9N_3O_2$) (methyl guanidine acetic acid or methyl-glycocyamine) has the following structural formula



From a study of the above formula it will be seen that from a chemical viewpoint creatine may be looked upon as a derivative of glycine, of guanidine or of guanidine acetic acid It is also related to arginine and probably to choline It is readily converted by acid *in vitro* to creatinine

Creatinine, $C_4H_7N_3O$ (methyl glycoyamidine) is the anhydride of creatine Its molecular constitution is shown by the following formula



Distribution of creatine and creatinine There are about 120 grams of *creatine* contained in the adult human body Of this 98 per cent is contained in the muscles and 1.5 per cent in the nervous system The remaining 0.5 per cent is distributed throughout the other organs of the body, of these, the testes contain it in highest concentration The skeletal and cardiac muscles and the gravid uterus are a great deal richer in creatine than the smooth muscle of the gastrointestinal tract and elsewhere About 80 per cent of the creatine in muscle is combined with phosphoric acid as *phosphocreatine* (ch 52) Of the striated muscles, the rapidly contracting, pale type contain more than the slowly contracting red variety (p 963) In the muscle of invertebrates creatine is replaced by arginine

Creatinine is present in muscle in much smaller amounts One hundred grams of skeletal muscle, for example, contain 300 to 500 mg of creatine

and only 10 mg of creatinine. The creatine of whole blood amounts to from 3.5 to 5.0 mg and the creatinine to about 2.5 mg per 100 cc. By far the greater part, if not all, of the creatine is contained in the corpuscles. The creatinine is about equally distributed between cells and plasma.

Excretion of creatine and creatinine. Creatine is normally absent from the urine of men, but is present in the urine of male and female children up to the age of puberty and frequently in the urine of women. The significance of these sex differences is not evident. In non-pregnant women the *creatinuria* is intermittent but is not related to menstruation. During pregnancy creatine is present continuously in the urine. This may be related to high creatine content of the uterus. For two or three weeks after child-birth the quantity excreted is even higher than during pregnancy. Reduction in the muscular mass of the uterus (involution) may account in part for the high output during the puerperium, but that uterine involution is not the chief cause is shown by the fact that it occurs although Caesarian section and removal of the uterus have been performed. Creatine also appears in the urine of either sex under the following conditions, *high protein diet, starvation, carbohydrate deprivation, diabetes, wasting diseases and fevers, exophthalmic goiter* and in certain *muscular dystrophies*. The *creatinuria* in all of these conditions except the last is probably due to an increase in the normal catabolism of muscular tissue, the liberation of creatine occurring more rapidly than its conversion to creatinine. Protein food, for example, has a stimulating effect upon metabolism (p. 641), while in wasting diseases, fevers, etc., tissue breakdown is accelerated. Carbohydrate deprivation probably acts indirectly in that its sparing effect upon protein catabolism is absent. In the muscular dystrophies, e.g., *myasthenia gravis, progressive muscular atrophy, amyotonia congenita, anterior poliomyelitis*, etc., the urinary creatine is probably derived from the degenerating muscle fibers, the muscles also appear to be defective in their power to store creatine. The *creatinuria* of normal children has received no satisfactory explanation, it is said to be due to an increased production of creatine, induced in some way by the growth impulse, and also probably to a low capacity for creatine storage of the undeveloped muscles. Another possibility is that children have a relatively low power to convert creatine to creatinine (Hunter).

The origin of creatine and creatinine. Creatinine is undoubtedly derived from creatine but there has been much uncertainty concerning the latter's origin. One or other of the substances to which its chemical formula suggests a relationship, e.g., guanidine, glycine or arginine, has been considered as a possible precursor. The opportunity afforded by the *creatinuria* of muscular dystrophies has been taken advantage of by Brand and his colleagues in the investigation of this question. Various amino-acids and other guanido-compounds were fed to patients suffering from such disorders, ingested glycine, or gelatin which is rich in this amino-acid, causes a pronounced rise (40 per cent) in the creatine excretion of these patients but not of normal persons. The proportion of ingested creatine which is retained is also much less than the normal (*creatinine tolerance test*). Arginine and glycoylamine (guanido-acetic acid) were also effective. Feeding benzoic acid, which drew upon the body's glycine store for detoxication purposes (p. 589) caused a decrease in the creatine excretion. The importance of glycine in creatine synthesis is emphasized by the observation of Almquist and Mecchi that glycine, which is an indispensable amino-acid for chicks, can be replaced in the diet of this species by creatine. These results pointed to arginine, guanido-acetic acid, and especially glycine as precursors of creatine.

The subject of creatine synthesis has been clarified recently through a number of experiments in which isotopes were used as markers to identify the particular substance under investigation after it had been absorbed. The administration of isotopic glycine (i.e., heavy nitrogen, N^{15} , incorporated in the glycine molecule) was shown by Bloch and Shoenheimer to give rise to creatine with a high concentration of the nitrogen isotope. Of all the other amino-acids investigated by this method, arginine alone was found to be a primary precursor of creatine. The isotope N^{15} was incorporated into the amidine part ($NH_2-C=NH$) of the arginine molecule. When the isotopic arginine was fed, the creatine isolated from the tissues of the animal was found to contain the isotope in this part of its molecule. The researches of du Vigneaud, and of Borsook and Dubnoff, show clearly that the methyl group of creatine is furnished by methionine. The former worker prepared methionine in which the three hydrogen atoms of the S-methyl group were present as the isotope H^2 (deuterium). When this was fed and

to 30 grams of urea) Minimal amounts are also excreted in the sweat (ch 53), salivary, intestinal and mammary secretions The urea-nitrogen varies with the protein content of the diet. Upon a low protein diet the output may be as low as 2 grams and on a diet rich in protein over 25 grams The value of urea-N is therefore taken as an index of the magnitude of the catabolism of food protein (exogenous metabolism, see table, p 637)

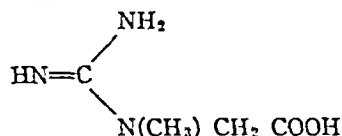
The blood contains from 8 to 15 mg of urea-N (18 to 35 mg urea) per 100 cc. Urea is readily diffusible and is found in about the same concentration in the various tissues and other fluids, e g, lymph, bile, cerebrospinal fluid and pancreatic juice, as in blood The kidney, however, is an exception since it contains some 150 to 200 mg of urea per 100 grams of tissue We have already seen that the urinary ammonia is formed from glutamine and amino acids in the kidney and not from urea (p 458)

THE SITE OF UREA PRODUCTION It has long been known that the liver is an important site of urea formation Perfusion of the liver with amino-acids or their addition to sliced liver tissue leads to the production of urea. Urea is very widely distributed throughout nature, being a constituent of the body fluids of the lowest forms of animal life. It is also found in plants For this reason its formation in animals might be thought to be a function common to all tissues But the experiments of Mann and his associates in hepatectomized dogs, points to the liver as being the main if not the sole site of urea production After complete removal of the liver the formation of urea ceased, so that its concentration in the blood and urine fell progressively If the kidneys as well as the liver were excised there was no accumulation of urea in the blood though the amino-acid concentration rose Also, after the injection of glycine or alanine into the hepatectomized animal, in contrast to the result obtained in the normal animal, the concentration of the injected amino-acid rose in the blood and urine, and there was no evidence of urea production That these results in dogs apply to man is indicated by the observation of Rabinowitch He has reported that in a case of extensive damage to kidneys and liver (acute yellow atrophy) the excretion of urine was almost suppressed yet the blood was urea free The appearance of leucine and tyrosine in the urine which occurs in acute yellow atrophy of the liver may also be ascribed to the failure of deamination of these amino-acids Never-

theless, though the liver undoubtedly plays the predominant rôle in the production of urea, certain other tissues may under certain circumstances share to some degree in its formation Graham and associates found, for example, that the active mammary gland produces not inconsiderable amounts

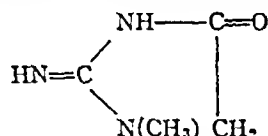
CREATINE AND CREATININE

Creatine ($C_4H_7N_3O_2$) (methyl guanidine acetic acid or methyl-glycococamine) has the following structural formula



From a study of the above formula it will be seen that from a chemical viewpoint creatine may be looked upon as a derivative of glycine, of guanidine or of guanidine acetic acid It is also related to arginine and probably to choline It is readily converted by acid *in vitro* to creatinine

Creatinine, $C_4H_7N_3O$ (methyl glycococamidine) is the anhydride of creatine Its molecular constitution is shown by the following formula



Distribution of creatine and creatinine There are about 120 grams of *creatine* contained in the adult human body Of this 98 per cent is contained in the muscles and 1.5 per cent in the nervous system The remaining 0.5 per cent is distributed throughout the other organs of the body, of these, the testes contain it in highest concentration The skeletal and cardiac muscles and the gravid uterus are a great deal richer in creatine than the smooth muscle of the gastrointestinal tract and elsewhere About 80 per cent of the creatine in muscle is combined with phosphoric acid as *phosphocreatine* (ch 52) Of the striated muscles, the rapidly contracting, pale type contain more than the slowly contracting red variety (p 963) In the muscle of invertebrates creatine is replaced by arginine

Creatinine is present in muscle in much smaller amounts One hundred grams of skeletal muscle, for example, contain 300 to 500 mg of creatine

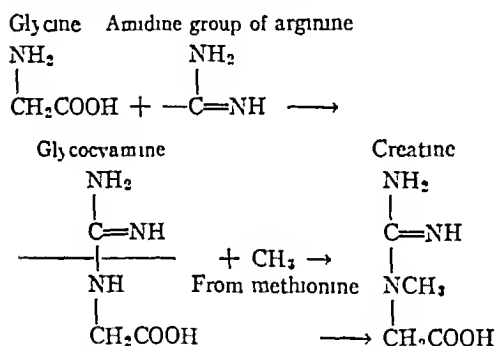
and only 10 mg of creatinine. The creatine of whole blood amounts to from 3.5 to 5.0 mg and the creatinine to about 2.5 mg per 100 cc. By far the greater part, if not all, of the creatine is contained in the corpuscles. The creatinine is about equally distributed between cells and plasma.

Excretion of creatine and creatinine. Creatine is normally absent from the urine of men, but is present in the urine of male and female children up to the age of puberty and frequently in the urine of women. The significance of these sex differences is not evident. In non-pregnant women the *creatinuria* is intermittent but is not related to menstruation. During pregnancy creatine is present continuously in the urine. This may be related to high creatine content of the uterus. For two or three weeks after child-birth the quantity excreted is even higher than during pregnancy. Reduction in the muscular mass of the uterus (involution) may account in part for the high output during the puerperium, but that uterine involution is not the chief cause is shown by the fact that it occurs although Caesarian section and removal of the uterus have been performed. Creatine also appears in the urine of either sex under the following conditions, *high protein diet, starvation, carbohydrate deprivation, diabetes, wasting diseases and fevers, exophthalmic goiter* and in certain *muscular dystrophies*. The *creatinuria* in all of these conditions except the last is probably due to an increase in the normal catabolism of muscular tissue, the liberation of creatine occurring more rapidly than its conversion to creatinine. Protein food, for example, has a stimulating effect upon metabolism (p. 641), while in wasting diseases, fevers, etc., tissue breakdown is accelerated. Carbohydrate deprivation probably acts indirectly in that its sparing effect upon protein catabolism is absent. In the muscular dystrophies, e.g., *myasthenia gravis, progressive muscular atrophy, amyotonia congenita, anterior poliomyelitis*, etc., the urinary creatine is probably derived from the degenerating muscle fibers, the muscles also appear to be defective in their power to store creatine. The *creatinuria* of normal children has received no satisfactory explanation, it is said to be due to an increased production of creatine, induced in some way by the growth impulse, and also probably to a low capacity for creatine storage of the undeveloped muscles. Another possibility is that children have a relatively low power to convert creatine to creatinine (Hunter).

The origin of creatine and creatinine. Creatinine is undoubtedly derived from creatine but there has been much uncertainty concerning the latter's origin. One or other of the substances to which its chemical formula suggests a relationship, e.g., guanidine, glycine or arginine, has been considered as a possible precursor. The opportunity afforded by the *creatinuria* of muscular dystrophies has been taken advantage of by Brand and his colleagues in the investigation of this question. Various amino-acids and other guanido-compounds were fed to patients suffering from such disorders, ingested glycine, or gelatin which is rich in this amino-acid, causes a pronounced rise (40 per cent) in the creatine excretion of these patients but not of normal persons. The proportion of ingested creatine which is retained is also much less than the normal (*creatinine tolerance test*). Arginine and glycocyamine (guanido-acetic acid) were also effective. Feeding benzoic acid, which drew upon the body's glycine store for detoxication purposes (p. 589) caused a decrease in the creatine excretion. The importance of glycine in creatine synthesis is emphasized by the observation of Almquist and Mecchi that glycine, which is an indispensable amino-acid for chicks, can be replaced in the diet of this species by creatine. These results pointed to arginine, guanido-acetic acid, and especially glycine as precursors of creatine.

The subject of creatine synthesis has been clarified recently through a number of experiments in which isotopes were used as markers to identify the particular substance under investigation after it had been absorbed. The administration of isotopic glycine (i.e., heavy nitrogen, N^{15} , incorporated in the glycine molecule) was shown by Bloch and Shoenheimer to give rise to creatine with a high concentration of the nitrogen isotope. Of all the other amino-acids investigated by this method, arginine alone was found to be a primary precursor of creatine. The isotope N^{15} was incorporated into the amidine part ($NH_2-C=NH$) of the arginine molecule. When the isotopic arginine was fed, the creatine isolated from the tissues of the animal was found to contain the isotope in this part of its molecule. The researches of du Vigneaud, and of Borsook and Dubnoff, show clearly that the methyl group of creatine is furnished by methionine. The former worker prepared methionine in which the three hydrogen atoms of the S-methyl group were present as the isotope H^2 (deuterium). When this was fed and

the creatine of the animals' bodies determined, the isotope concentration indicated that nearly 70 per cent of the methyl groups of the creatine molecule had been derived from methionine. Borsook and Dubnoff incubated liver slices with glycocyamine acid (glycocyamine) which they found was slowly converted to creatine, the conversion being hastened by the addition of methionine. Glycocyamine is apparently an intermediary in the synthesis of creatine in the body, when the isotopic form of this compound is fed isotopic creatine is formed. The conclusions to be drawn from these experiments are that arginine, glycine and methionine enter into the synthesis of creatine but that other amino-acids do not serve as precursors. The first reaction, it appears, is between the amidine part of arginine and glycine with the formation of glycocyamine, arginine furnishing the amidine group of the latter. The guanidoacetic acid then undergoes methylation by the transference of the methyl group from methionine. Thus



The site of creatine production is not known definitely but it is probably in the muscles. These also in all probability are the site of creatine-creatinine transformation (by dehydration). Mann and Magath found that removal of the liver was without effect upon creatinine production. The creatine which gives rise to urinary creatinine is derived very largely from the phosphocreatine of muscle.

Creatine is not, as was held by some authorities in the past, simply a waste product of protein metabolism like urea. If, for example, a large dose of creatine is fed the greater part or the whole is retained in the body.² Creatinine, on the other

hand is purely a waste product. Up to 80 per cent or more of an amount fed can be recovered unchanged in the urine. The investigations of recent years into the chemistry of muscular contraction have revealed the essential importance of creatine (as phosphocreatine) in the contractile process (ch 52).

Creatinine excretion as an index of muscle metabolism

The daily output of creatinine in the urine is constant for the individual, amounting to from 1.5 to 2.0 grams for men and from 0.8 to 1.5 grams for women. This corresponds to about 2 per cent of the creatine of the body. Unlike the excretion of urea, which is derived largely from exogenous sources, the creatinine output is practically independent of the protein level of the food. This is evident from table 55. The creatinine excretion is therefore considered to be an index of the magnitude of the metabolism of the tissues and especially of muscle. The daily output of creatinine is extraordinarily constant for the individual; it is not influenced by ordinary exercise or by the urine volume. The creatinine coefficient—

$$\frac{\text{milligrams creatinine}^4 \text{ excreted per day}}{\text{body weight in kilograms}}$$

is from 20 to 26 for the majority of normal men and from 14 to 22 for women. Its value depends upon the muscular development of the individual, the sex variation being due presumably to the different relative amounts of fatty and muscular tissues of male and female bodies. Athletic women for this reason have a coefficient as high as or higher than a man of obese build and poor muscular development.

THE METABOLISM OF SULFUR

Sulfur enters the body mainly as a constituent of the amino acids *cystine* and *methionine*. Food also contains inorganic sulfates, e.g., sodium and potassium sulfates, small amounts of sulfur in the form of sulfo-lipids (sulfatides), and sulfur combined in certain glycoproteins as mucortin-sulfuric

when the "creatinine reservoirs" had apparently become filled a large part was excreted unchanged. A rise in the creatinine output also occurred, indicating the transformation of creatine to creatinine. Nevertheless, of some 33 grams of creatine administered over the entire period of 70 days about 20 grams were retained. Retention of ingested creatine is also demonstrable in man.

⁴ Or creatine + creatinine, when creatinuria exists.

² In the experiments of S. R. Benedict and Osterberg over half a gram of creatine hydrate was fed daily to dogs for a period of several weeks. In the first week none of the administered creatine was excreted, either as such or as creatinine. During subsequent weeks

acid and chondroitin-sulfuric acid Sulfur in inorganic form cannot be used in the construction of body protein Practically speaking, the body is dependent for its sulfur supplies upon the two sulfur-containing amino-acids mentioned above Food sulfur may be summarized as follows

A Organic

(1) Protein sulfur (major source)

- (a) Sulfur-containing amino-acids, cystine and methionine
- (b) Glycoproteins, as mucoitin-sulfuric acid in mucin and ovomucoid, and chondroitin-sulfuric acid in chondromucoid of cartilage (p 623)

(2) Non-protein sulfur

Sulfo-lipids

B Inorganic

Potassium, sodium and magnesium sulfates

THE DISTRIBUTION OF SULFUR IN THE BODY

Sulfur is contained in the ordinary tissue proteins, in hair, horn, feathers, etc., in mucin, as mucoitin-sulfuric acid and chondroitin-sulfuric acid, in certain glycoproteins of tendons, vitreous humor, cornea and connective tissues, in glutathione and insulin, in the taurocholic acid of bile, as sulfocyanate in saliva, in ergothionine, a compound found in red corpuscles, in certain pigments (melanins, urochrome), and in nervous tissue as

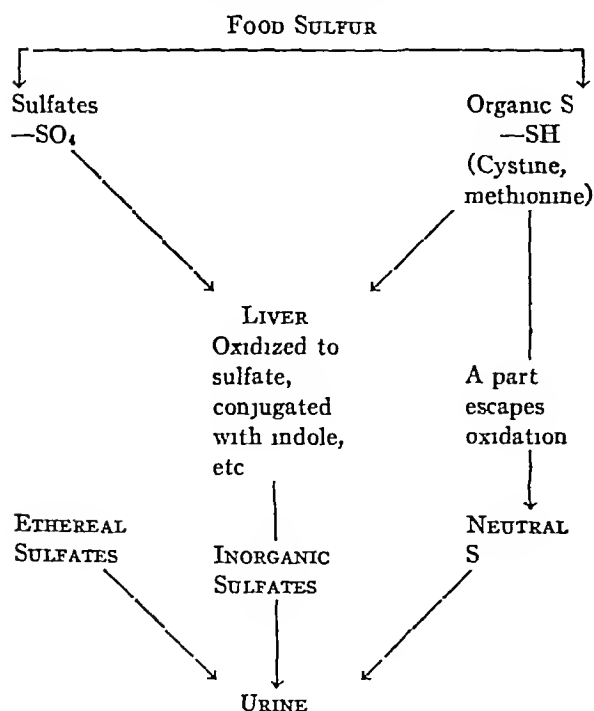


TABLE 55
(After Folin)

	NITROGEN RICH DIET	NITROGEN POOR DIET
Volume of urine	1170 cc	385 cc
Total nitrogen	16.8 grams	3.60 grams
Urea nitrogen	14.70 grams = 87.5%	2.20 grams = 61.7%
Ammonia nitrogen	0.49 gram = 3.0%	0.42 gram = 11.3%
Uric acid nitrogen	0.18 gram = 1.1%	0.09 gram = 2.5%
Creatinine nitrogen	0.58 gram = 3.6%	0.60 gram = 17.2%
Undetermined ni- trogen	0.85 gram = 4.9%	0.27 gram = 7.3%
Total sulphur	3.64 grams	0.76 gram
Inorganic SO ₂	3.27 grams = 90.0%	0.46 gram = 60.5%
Ethereal SO ₂	0.19 gram = 5.2%	0.10 gram = 13.2%
Neutral S	0.18 gram = 4.8%	0.20 gram = 26.3%

sulfolipoids Inorganic sulfates are contained in the body fluids generally

The loss of sulfur from the body occurs in the shedding of hair, nails, etc., in the bile, saliva and gastro-intestinal secretions The great bulk of the sulfur loss, however, occurs through the kidneys

The history of sulfur in the body The sulfur liberated in the catabolism of dietary protein is largely converted to inorganic sulfates A part of the inorganic sulfate derived from this and other sources becomes conjugated in the liver with substances produced in the intestine by the bacterial decomposition of protein to form *etheral sulfates* (p 589) The latter are excreted in the urine The products of bacterial action, among which are phenol derived from phenylalanine and tyrosine, and indole and skatole from tryptophane, possess toxic properties Their excretion as etheral sulfates constitutes a detoxicating mechanism When the detoxicating function of the liver is depressed as a result of hepatic disease, some of these toxic substances are excreted by the kidney in abnormal amounts in the free state

Urinary sulfur The total urinary sulfur is made up of the following

- (1) Inorganic sulfur (85 to 90 per cent) Compounds of sulfuric acid with Na, K, Ca and NH₃

(2) Ethereal sulfate sulfur (6 to 8 per cent), e.g., potassium and sodium salts of indoxyl sulfuric acid (p 590) The former is known as indican

(3) Neutral sulfur (4 to 6 per cent) e.g., sulfur-containing amino-acids, urochrome, thiosulfates, thiocyanates, taurocholic and oxyproteic acids

The quantities of inorganic and ethereal sulfates in the urine vary with the protein level of the diet, and run more or less parallel with the nitrogen excretion, the neutral sulfur is influenced to a less degree. For this reason the two former partitions of the urinary sulfur were considered by Folin to represent food (exogenous) sulfur while the neutral sulfur was taken to be derived mainly from body protein (endogenous sulfur) (see p 632). On a diet of meat or during prolonged starvation the ratio of sulfur to nitrogen in the urine is about 1 to 14, i.e., for every gram of sulfur there are approximately 14 grams of nitrogen. This is approximately the S/N ratio found in muscular tissue. A graphic summary of sulfur metabolism is shown on page 637.

As a result of the excessive production and absorption of putrefactive products the excretion of ethereal sulfates is increased in acute intestinal obstruction. They are also increased in carcinoma of the liver. Chronic constipation, however, exerts little or no influence upon the excretion of these substances (see p 589-90).

The excretion of neutral sulfur is increased in the rare metabolic anomaly known as *cystinuria* and in cases of *melanotic sarcoma*, when an abnormal sulfur-containing pigment appears in the urine.

The non protein sulfur of blood. The sulfur of blood, other than that present as a constituent of protein, amounts to from 3 to 5 mg per 100 cc. The concentrations of the three forms are given in table 56.

AMMONIA (SEE PAGE 458)

NITROGEN BALANCES

The difference between the nitrogen taken in the food and that excreted in the feces and urine is spoken of as the nitrogen balance (that lost in the hair, sweat, saliva, etc., is negligible). On an ordinary diet, the amounts of nitrogen excreted daily in the feces and urine, respectively, average 1.5 and 13 grams. The nitrogen of the feces is in part unabsorbed nitrogen of the food and in part nitrogen excreted through the intestinal wall. When the intake and output are equal, the body

is said to be in *nitrogen equilibrium*. When the intake exceeds the output the body is in *positive nitrogen balance*—nitrogen is being retained. If the reverse is the case, i.e., the output exceeds the intake, the balance is said to be *negative*—the body is losing nitrogen. When the nitrogen of the food is increased a greater quantity for a time thereafter is retained, but the balance soon becomes adjusted to the higher intake and gradually returns to its original value (table 57).

The healthy adult requires protein to replace the inevitable loss of tissue protein. When given a sufficient amount of protein to repair this so-called "wear and tear" his output and intake of nitrogen will balance, i.e., he will be in nitrogen equilibrium. If his diet contains protein in excess of this amount the nitrogen not employed for repair is excreted and nitrogen equilibrium maintained. In starvation or on a low protein intake, or on a diet which lacks certain essential amino-acids, the individual continues to excrete nitrogen derived from the dissolution of his own protoplasm, and so goes into negative balance. In children, in adults recovering from wasting diseases or undergoing muscular training, and in pregnant women, the body, if the protein intake is liberal, does not excrete as much nitrogen as it received. Nitrogen is retained for the manufacture of new tissue. The daily excretion of nitrogen of a man of average weight upon a nitrogen-free diet amounts from 0.75 to 3 grams. It might be thought that nitrogen equilibrium could be reestablished by simply feeding this quantity of protein nitrogen. It is found however that a considerably larger quantity is necessary. The reason for this is that the amino acids are required by the body in different proportions from those present in the food protein. In order that the body shall obtain a sufficient quantity of suitable amino acids for the replacement of its own protoplasm and for the manufacture of various secretions and other essential materials, a relatively large assortment must be available from which it can choose. Those not required are discarded. For this reason, equilibrium can be established upon a smaller quantity of animal protein, such as beef muscle which contains amino-acids more nearly in the same proportions as those in human body protein, rather than upon vegetable proteins. The wastage of nitrogen is less in the former instance. If the only protein in the food is one entirely lacking in certain essential amino acids (lysine, histidine, etc.) the loss of

nitrogen resulting from the breakdown of such amino-acids in the body's protoplasm cannot be replaced. A negative balance results no matter how much of the inferior protein has been fed (see ch 47)

The quantity of protein required to establish nitrogen equilibrium depends very greatly upon the content of the diet in the other two food principles—fats and carbohydrates. It is impossible, for example, to establish nitrogen equilibrium in man upon an exclusively protein diet, the excretion of nitrogen always exceeds the intake even though the individual ingests protein to his full capacity. The reason for this is that man cannot consume and digest sufficient amounts of protein to satisfy his energy requirements. Under such circumstances he draws upon his stores of carbohydrates and fats, but after these have been exhausted the protein elements of his tissues are disrupted. The non-nitrogenous portions of the amino-acid molecules are burned to make up the calorie deficiency, and, as a consequence, quantities of nitrogen derived from food and body protein are excreted. It has been calculated that the human subject would have to consume, daily, some 8 pounds of meat—practically an impossible feat for a civilized man—in order to furnish the necessary energy and maintain the body in nitrogen equilibrium. On the contrary, a carnivorous animal such as the dog which possesses a large capacity for the digestion of protein food⁶ thrives upon a diet composed entirely of lean meat. The Eskimos are also capable—according to Krogh—of consuming relatively enormous quantities of meat. On a mixed diet, as we shall see immediately, nitrogen equilibrium can be established on a very low protein intake.

PROTEIN SPARERS

Carbohydrates and fats are called *protein spacers* since their presence in the diet relieves tissue protein of the necessity of furnishing energy. The protection of body protein by carbohydrate has been shown clearly by numerous experiments. The following experiment performed by Lusk upon himself may be cited in illustration of this impor-

⁶ The daily energy requirement of man is, say, 3000 Calories. Eight pounds of meat (3600 grams approximately) which is 20 per cent protein, has a caloric value of $(20/100) \times 3600 \times 4.1 = 2916$. The energy expenditure of an average sized dog, on the other hand, is around 600 Calories. It can consume in less than a minute 2 pounds (900 grams) of meat, yielding more than 700 Calories.

TABLE 56
Concentrations of different forms of sulfur in whole blood and serum
(After Denis)

	WHOLE BLOOD	SERUM
	per 100 cc mg	per 100 cc mg
Inorganic sulfur	0.1-1.1	0.5-1.1
Ethereal sulfate sulfur	0.1-1.0	0.1-1.0
Neutral sulfur	2.2-4.5	1.7-3.5

The inorganic sulfur of blood is elevated in *renal insufficiency, intestinal obstruction and leukemia*.

tant principle. After a period during which a mixed diet had been fed and a small positive nitrogen balance established, 350 grams of carbohydrate were withdrawn from the diet. The nitrogen excretion increased from 19.84 to 27.00 grams daily, the balance becoming negative. Also, the nitrogen excretion in the urine in starvation is nearly 3 times greater than that upon a diet, nitrogen-free, but of adequate energy value (i.e., one composed exclusively of fat and carbohydrate). The non-protein food has reduced the break-down of tissue protein.

The nitrogen excretion on such a protein-sparing diet is therefore an index of the inevitable disintegration of body protoplasm which occurs under ordinary physiological conditions. For this reason it was called, by Rubner, the "wear and tear" quota of protein metabolism (see p. 631). One of the lowest values reported for nitrogen excretion is that of Boothby and Sandiford, namely, 1.74 grams daily (0.024 gram per kilogram of body weight).

TABLE 57
Example of adjustment of nitrogen balance to increased intake

DAY	NITROGEN IN FOOD	NITROGEN IN FECES	NITROGEN "ABSORBED"	NITROGEN IN URINE	NITROGEN BALANCE
	grams	grams	grams	grams	grams
1	14.40	0.70	13.70	13.60	+0.10
2	14.40	0.70	13.70	13.80	-0.10
3	14.40	0.70	13.70	13.60	+0.10
4	20.96	0.82	20.14	16.80	+3.34
5	20.96	0.82	20.14	18.20	+1.94
6	20.96	0.82	20.14	19.50	+0.64
7	20.96	0.82	20.14	20.00	+0.14

* Modified from H. C. Sherman, "Chemistry of Food and Nutrition" 4th edition.

Carbohydrate exerts a specific sparing action quite apart from the fact that it furnishes energy and so relieves protein of the necessity of performing this duty, for the sparing effect exhibited by a given quantity of carbohydrate cannot be brought about by an amount of fat possessing double the caloric value. Carbohydrate when fed alone has a marked sparing action, whereas fat alone has little or no protein-sparing effect and a positive nitrogen balance cannot be established in man on a diet composed entirely of fat and protein. Nor will fat by itself, when given after a period of fasting, reduce the nitrogen excretion below that during the fast. Indeed, it actually increases the output of nitrogen. A diet whose calories, up to 50 per cent or more, are derived from fat and the remainder from carbohydrate has, however, as great a protein-sparing action as one whose calories are derived from carbohydrate alone. Lactic and pyruvic acids, products of carbohydrate metabolism, are also protein-sparers to some extent. The sparing effect of carbohydrates can be demonstrated in the well-nourished animal as well as during starvation. When dogs in nitrogen equilibrium are given extra glucose, nitrogen is retained but the balance is re-established when the administration of glucose is stopped (Larson and Chaikoff).

TABLE 58

The results of an experiment of Rapport's which illustrates the utilisation of the "waste heat" of the specific dynamic action of fat and glucose in exercise

Heat production at rest and in exercise and recovery

	REST ING (PER HOUR)	DEVI- ATION FROM (A)	EXCESS IN EX- ERCISE AND RECOV- ERY (PER KGM. OF WORK)*	DEVI- ATION FROM (A)
	Calo- ries	per cent	Calo- ries	per cent
A In the post absorp- tive state	12.7		2.40	
B After ingestion of fat	15.2	+19.7	2.54	+5.8
C After ingestion of glu- cose	15.4	+21.3	3.8	-0.8
D After ingestion of meat				
100 grams	16.2	+27.6	3.38	+40.8
200 grams	17.8	+40.1	3.48	+45.0

* Based on $\frac{1}{2}$ hour period after beginning of exercise.

The difference between the sparing actions of fat and carbohydrate is not easily explained. According to one view (Landergren) the specific sparing effect of carbohydrate depends upon the fact that glucose is vitally essential to the body and must be constantly supplied whatever the cost. If it is not available from outside sources or from the glycogen stores, tissue protein is broken down to yield materials from which it may be synthesized.

Another more probable explanation of the peculiar sparing action of carbohydrate is, that it is used in the synthesis of the amino-acids which are incorporated into body proteins. According to this view the nitrogen resulting from the breakdown of tissue protein during fasting (or from tissue and food protein when the diet consists exclusively of protein and fat) is largely or entirely excreted. If, on the other hand, certain intermediary products of carbohydrate metabolism are available, e.g., pyruvic acid, etc., it is possible for amino acids to be resynthesized. That is, nitrogen which would otherwise be wasted is retained.

THE SPECIFIC DYNAMIC ACTION OF FOOD

The heat production of a subject under basal conditions, except that he has recently ingested food, is raised considerably above the level in a truly basal state. This calorigenic effect of food is called its specific dynamic action (SDA). The heat production commences to rise within an hour after the food has been eaten, attains a maximum in about the third hour and is maintained above the basal level for several hours. The greatest specific dynamic action is exerted by protein food. When protein is fed alone to a fasting animal, in an amount possessing a heat value equivalent to the animal's estimated basal metabolism, the heat production is raised 30 per cent or more above the basal level. Carbohydrate causes a rise of about 6 per cent and fat of 4 per cent. That is to say, when a quantity of protein, carbohydrate or fat possessing an energy value of 100 Calories is fed separately to an animal whose basal metabolism is 100 Calories daily, its actual heat production will be 130, 106 or 104 Calories, respectively. The extra heat is generated by the combustion of body substance, so if loss of weight is to be prevented such an animal must be supplied with 130 Calories if his diet is protein, or if carbohydrates and fats are also consumed extra allowances must be made

for the amounts of these substances which have been included

To take a theoretical case in illustration, if a man with a basal metabolism of 1500 Cal were kept on a 1500 Cal diet, his body weight would decline until the total heat production (basal + SDA) would just equal the energy value of the food, from there on, the body weight, and total metabolism would remain steady at the new levels. But if the energy value of the food were again reduced to the point where it just balanced the basal metabolism (no allowance being made for SDA) a further weight loss and diminution in basal metabolism would occur, and should the adjustment of food to basal heat production be carried out repeatedly, the body tissues would be gradually consumed, as long as the subject survived, in order to furnish the extra heat caused by SDA. In planning a diet, therefore, an extra caloric allowance must be made for the SDA of the food itself (ch 56)

The extra heat resulting from protein food cannot be employed for the production of mechanical or other forms of energy. It is waste heat and is simply added to heat produced by the muscular exertion. A diet very rich in protein is therefore unsuited to heavy muscular work. The SDA of protein is an important factor, however, in the regulation of body temperature (p 736). With fat and carbohydrate the case is different, for the extra heat is harnessed in the performance of work. When exercise is undertaken upon either of these substances the heat due to their specific dynamic action is almost abolished, the extra energy being incorporated in the energy exchanges of the exercise (table 58)

When new tissue is being formed, i.e., when the nitrogen balance is positive, protein does not exert its usual specific dynamic action. Nor does it occur in a fever, such as typhoid, when there is great destruction of tissue protein. The ingested food then merely replaces or spares the tissue protein.

The cause of the specific dynamic action

Several explanations for the phenomenon have been offered. It is certainly not due to digestive processes, i.e., to the contractions of smooth muscle of the alimentary canal or to the work entailed in the secretion of the digestive juices. This is proved by the following facts. Bones given to a dog, or agar agar, saline cathartics, water or meat extracts (which stimulate powerfully the gastric secretion) given to man, have no effect upon the

heat production. Also, as already mentioned, protein has no SDA when new tissue is being laid down. On the other hand, the injection of certain amino-acids into the blood stream is followed by a specific dynamic effect.

The SDA of protein has been shown to depend upon the following seven amino-acids—*glycine, alanine, leucine, glutamic* and *aspartic acids, histidine* and *tyrosine*. Tyrosine has the greatest effect, glutamic acid comes second. The specific dynamic effect of glycine is greater than that of alanine but either of these exerts a smaller effect than tyrosine, the effect of leucine is slight. The SDA of a given quantity of meat equals that calculated from its content in these amino-acids. The other amino-acids investigated have no specific dynamic action.

The cause of the SDA is still imperfectly understood though it is generally accepted that the extra heat produced by *protein* food is associated with the metabolism of the amino acids. They exert in some way a stimulating effect upon the tissue cells, raising their heat production to a high level, i.e., the oxidation of the cells' own fuel material is increased, the rate of other metabolic processes within the cells is raised. The extra heat is not due to the amino-acids themselves being utilized as fuel nor to a stimulating action upon the tissue cells of these materials in the unchanged state. It is in the intermediary reactions of the amino-acids that the specific dynamic action of protein should be sought—e.g., in the deamination process. Borsook and Winegarden conclude from their studies that the SDA of protein results from the metabolism (deamination and urea production) and the excretion of nitrogen, an increased excretion of nitrogen accompanies the greater heat production. According to Borsook an increase of from 7 to 10 Calories of extra heat are produced for each gram of extra nitrogen excreted. The remainder of the extra heat produced is a more variable and usually a larger fraction of the total, it is attributed to the metabolism of the carbon part of the protein molecule. There is a considerable body of evidence for the view that the deamination of the amino-acids with the formation of urea contributes very considerably to the specific dynamic action of protein. Lundsgaard found, for example, that though the administration of sodium acetate or sodium lactate was followed by only a slight increase in heat production, ammonium acetate or ammonium lactate caused a pronounced rise, even ammonium chloride caused a well marked increase in heat production.

The reactions responsible for the SDA of protein are apparently situated in the liver and not in the tissue cells generally, since Wilhelmj, Bollman and Mann were unable to obtain any effect in hepatectomized

dogs following the injection of amino-acids. Dock also found that after the administration of casein to rats the oxygen consumption of the hindquarters (muscular tissue) was 8 per cent greater than in the corresponding tissues of a control group, whereas the oxygen consumption of the abdominal viscera was 141 per cent greater

The specific dynamic action of carbohydrate is thought to represent the energy liberated in excess of that required for the conversion of glucose to glycogen. After a fast which depletes the glycogen stores, ingested glucose is oxidized in negligible amounts, yet it causes a pronounced specific dynamic action (Dann and Chambers)

The specific dynamic action of fat is ascribed to the increased concentration of fat in the tissue fluids and, as a consequence, to its more rapid oxidation ("plethora theory" of Lusk)

The glands of internal secretion appear to have no direct influence upon the SDA of protein, the usual effect has been observed in a cretin with a basal metabolic rate 20 per cent below normal and it is not altered in hyperthyroidism. Thyroidectomy in animals is said, however, to reduce the SDA of carbohydrate and fat. Cushing and Fulton found the SDA of protein within the normal range in a number of cases of pituitary disease (hypopituitarism and acromegaly). Gaebler found it of normal value in hypophysectomized animals. In undernutrition the SDA of all foods is increased. It is, according to some observers, diminished in simple obesity (p. 710)

The SDA of carbohydrate and protein or of carbohydrate and fat is less than the sum of the values of each when fed separately. There is only a slight discrepancy, however, between the values for protein and fat, fed together or separately.

CHAPTER 47

PROTEIN METABOLISM (*Continued*)

THE NUTRITIONAL VALUE OF VARIOUS PROTEINS

Two factors must be considered in determining the nutritional value of a given protein (a) its digestibility, i.e., the amount absorbed, and (b) its suitability for the construction of tissue protein

(a) **THE DIGESTIBILITY OF PROTEINS** Though by far the greater part of the nitrogen arising from the catabolism of amino-acids within the body is excreted in the urine, a small fraction is eliminated through the secretions of the digestive glands and intestinal mucosa. This, the so-called *metabolic nitrogen*, is estimated from the nitrogen of the feces upon a nitrogen-free diet. In a man it amounts to from 0.5 to 1.5 gram daily, varying in amount with the bulk (roughage) of the diet. In order, therefore, to determine the proportion of a given protein which has undergone digestion and absorption, the total nitrogen content of the ingested protein¹ is determined (each 6.25 gm. of protein equals 1 gm. of nitrogen), and from this value the nitrogen of the feces less that for the metabolic nitrogen is subtracted. Thus—

$$\text{food N} - (\text{feces N} - \text{metabolic N}) = \text{absorbed N}$$

The digestibility of the particular protein is then expressed as a percentage of the food nitrogen which has been absorbed. This percentage value is referred to as the *coefficient of digestibility*. For example, if the food contains, let us say, 10 grams of nitrogen and it is found that 9.5 grams have been absorbed the digestibility is 95 per cent. Digestibility in the sense just defined is quite different from the popular meaning of the term which refers rather to the subjective sensations accompanying digestion.

Proteins of animal origin have the highest digestibility which runs from 95 to 100 per cent. That is, the wastage in digestion is 5 per cent or

¹ The total N in a sample of protein food is somewhat more than that actually present as protein-N. That is, a small part is present in the form of free amino-acids, their derivatives and other non-protein substances. Since much of the nitrogen in these forms is believed to be of definite nutritive value, no greater error is entailed in making the calculations upon the basis of the *total* food nitrogen than if a correction were made for their presence.

less. The digestibility of the proteins of nuts and fruits is low, that of the proteins of legumes and of potatoes is around 80 per cent (see table 59). The digestibility of wheat proteins is from 90 to 100 per cent thus approaching that for animal proteins.

(b) **THE SUITABILITY OF A GIVEN PROTEIN FOR THE SYNTHESIS OF BODY PROTEIN** This depends upon the amino-acid constitution of the protein. It is obvious that the greater the proportion of amino-acids in the dietary protein which can serve for the construction of tissue protein, the greater will be its potential nutritive value. In other words, the more closely the amino-acid assortment in the food protein resembles that in body protein the less of the former need be furnished. A smaller proportion of the amino-acids will then be discarded to have their nitrogen eventually being excreted in the urine, i.e., less of the food protein will be wasted.

Methods employed for estimating the nutritive values of proteins Osborne and Mendel, employing rats, added a known amount of the protein to be tested to a diet free from all other proteins, but adequate in other respects (i.e., one possessing the necessary energy value, minerals, and the vitamins then known). They expressed the nutritive value of the protein as the weight, in grams, gained by the animals per gram of protein fed, in the case of young rats (value for growth), or in the case of adult rats, the smallest quantity of the fed protein in grams, per gram of rat per week, necessary to maintain a constant body weight (value for maintenance). McCollum and associates also employed rats, feeding a basal diet of first class quality in all respects except that it lacked protein. The lack was made good by adding the protein to be tested, and the percentage of protein required to be added to the food mixture in order to promote normal nutrition gave an index of its nutritive value. The general state of nutrition of the animals, as indicated by rate of growth, fertility, care of young and longevity, was noted over a period corresponding to two-thirds or more of their life-span. On this basis, proteins were classed as "excellent" "good" or "poor". An excellent or first-class protein is one which "will support nearly optimum nutrition over periods approximating two-thirds or more of the normal life span of the rat when fed in amount corresponding to 9 per cent in the food-mixture." Nine per

TABLE 59

*Protein values of foods for maintenance and growth
Level of protein feeding, 8 to 10 per cent
(After Mitchell and Hamilton)*

FOOD	WATER CON- TENT*	PRO- TEIN CON- TENT* ON FRESH BASIS	QUALITY OF PROTEIN		META- BOLIC PROTEIN IN FECES† ON FRESH FOOD BASIS	PRO- TEIN VALUE OF FOOD ON FRESH BASIS
			Diges- ti- bility	Bio- logical value		
	per cent	per cent	per cent	per cent	per cent	per cent
Whole egg‡	73 2	13 2	100	94	0 4	12 0
Milk	87 0	3 3	100	85	0 2	2 6
Egg white‡	86 2	12 3	100	83	0 2	10 0
Beef liver	71 2	20 4	90	77	0 4	14 9
Beef kidney	76 7	16 6	99	77	0 3	12 3
Beef heart	62 6	16 0	100	74	0 5	11 3
Beef round	70 0	21 3	96	69	0 4	13 7
Pork ham	60 0	25 0	100	74	0 6	17 9
Veal§	73 4	20 7	100	62	0 4	12 4
Rolled oats	7 7	16 7	90	65	1 3	9 8
Whole wheat	11 4	13 8	91	67	1 3	7 1
White flour	12 8	10 8	100	52	1 3	4 3
Whole corn	10 3	7 5	95	60	1 3	3 0
Potato	78 3	2 2	78	67	0 3	0 8
Navy beans‡	12 6	22 5	76	38	1 3	4 2

* Average analyses taken, as far as possible, from Bull 28 (revised), Office of Experiment Stations, U S Dept Agr

† The metabolic nitrogen in the feces is assumed to equal 0.23 gram per 100 grams of dry matter of food. See Bull Natl Research Council, 1926, *x*, pt. 1, no 55, p 23

‡ Cooked

§ The cut tested was not recorded. It proved to be very fibrous. Analysis for shoulder cut assumed

cent is therefore the critical level for a protein of the highest nutritive value. Inferior proteins must be fed at higher levels. The relative values of several foods are shown below. The proteins of beef liver and kidney head the list, the others are shown in descending order of values

Beef liver and kidney
Egg (whole)
Beef muscle, fish
Milk
Soya bean
Oats (rolled)
Wheat (whole)

The determination of the minimal quantities of different proteins required for the maintenance of nitrogenous equilibrium in adult animals or of a positive balance (storage of nitrogen) in growing animals is also

used as a means of arriving at the nutritional values of proteins

The nutritive value of a particular protein, based upon the nitrogen metabolism following the feeding of known amounts, is employed by Mitchell. In this method, the value of the protein for repair (maintenance) or building of new tissue (growth) is given as a percentage of its *absorbed* nitrogen which is not lost in the urine, i.e., the percentage retained in the body. This is called its *biological value*. The animal is placed upon a basal diet complete in so far as non-protein factors are concerned. In order that none of the protein shall be burned for energy purposes, its level in the diet is kept low, i.e., there should be no excess over that required for growth or maintenance. The following example from Mitchell will serve for illustration.

A rat receiving a diet containing about 4 per cent of protein, ingested daily 56.9 mg of nitrogen and excreted 27.6 mg of nitrogen in the feces. Of the latter 21.7 mg constituted metabolic nitrogen (p. 643). The unabsorbed N was therefore only (27.6 - 21.7 =) 5.9 mg, and the absorbed N (56.9 - 5.9 =) 51.0 mg. The daily urinary N was 48.6 mg. Of this, 37.7 was derived from the body tissues (estimated from urinary N on a nitrogen free diet (ch. 46)). The total urinary nitrogen less this value for the endogenous nitrogen must represent the quantity of absorbed nitrogen which was excreted, that is, (48.6 - 37.7 =) 10.9 mg is the quantity of absorbed nitrogen which had been wasted in metabolism. The absorbed nitrogen, as just stated, amounted to 51.0. So, (51.0 - 10.9 =) 40.1 mg were retained in the body. The biological value of the dietary protein was therefore ((40.1/51.0) × 100 =) 79 per cent.

In order to know the value of a food as a source of suitable protein the quantity of protein which it contains as well as the latter's biological value must be considered. Thus a food may contain a poor protein in larger amounts or a first-class protein in small amounts. Again, in some foods the protein is both poor in quality and low in amount while in others both the quantity and quality of the protein are high. So we may take these two factors into consideration and speak of the *protein value* of a food. The quantity of protein in a sample of food is obtained by multiplying the value for its nitrogen content by 6.25 since pure animal food protein contains on an average 16 per cent of nitrogen. The use of this factor assumes, of course, that all the nitrogen in the food is protein nitrogen but, as mentioned in the footnote on page 643 a variable amount is present as non-protein nitrogen.

The values of individual amino acids were tested by Osborne and Mendel by feeding purified proteins known to be deficient in their amino acid constituents. The weights of the animals on the deficient protein were observed for periods before and after the missing amino acid had been added to the diet. Rose's method consists in feeding a mixture of purified amino acids

such as are found in casein. The effect upon growth was then determined after one or other amino acid had been omitted from the mixture.

The value of a protein in nutrition depends entirely upon its amino-acid constitution. Block and Mitchell have shown a very close correlation between the biological value of a protein and its amino-acid content. Thus, by the hydrolysis of a protein and determining the percentages of the various amino groups of which it is composed, a very accurate appraisal of its nutritive value can be made. Though animal proteins are usually superior to those of plant origin, no generalization can be made in this respect. It is a matter of the amino-acids of which the protein, animal or vegetable, is made up. In table 60 the amino-acid percentages in the proteins of liver and of soy bean are compared.

THE ESSENTIAL IMPORTANCE OF CERTAIN AMINO-ACIDS FOR GROWTH AND MAINTENANCE

In 1907 Willcock and Hopkins found that when young mice were fed upon a diet which contained zein (a protein of maize) as its sole protein, growth was arrested and the animals died in about 17 days. Zein is almost completely free from tryptophane. Yet, the addition of this amino-acid to the diet was not capable of promoting growth, and the survival period of the animals was extended to only 33 days. The addition of tyrosine was without effect since zein already contains this amino-acid in adequate amounts. It was shown by Osborne and Mendel that when lysine, which is also absent from this protein, was added to the zein diet together with tryptophane, the animals grew normally and remained in good health (fig 47 1). Lysine added alone was of no more benefit than tryptophane alone. Gelatin, like zein, is an incomplete protein. It lacks tyrosine and tryptophane and is very deficient in cystine. A diet which contains gelatin as its sole protein will not permit growth nor even maintain nitrogen equilibrium in an adult animal. When, however, the lacking amino-acids are added to the gelatin diet its defects are corrected.

Casein. Casein is deficient in glycine and has a low cystine content. When, however, casein is fed in sufficient quantity—about 18 per cent of the diet—or if cystine be added, normal growth results (fig 47 2).

Edestin (a protein in hemp seed) is relatively poor in lysine as is also *gladin* (of wheat) while

TABLE 60
(From data of Block and Mitchell)
Approximate amino-acid content of liver and of soy bean compared
(Calculated to 16.0 g of nitrogen)

AMINO-ACID	LIVER	SOY BEAN
	g	g
Arginine	6.6	7.1
Histidine	3.1	2.3
Lysine	6.7	5.8
Tyrosine	4.6	4.1
Tryptophane	1.4	1.2
Phenylalanine	6.1	5.7
Cystine	1.4	1.9
Methionine	3.2	2.0
Threonine	4.8	4.0
Leucine	8.4	6.6
Isoleucine	5.6	4.7
Valine	6.2	4.2

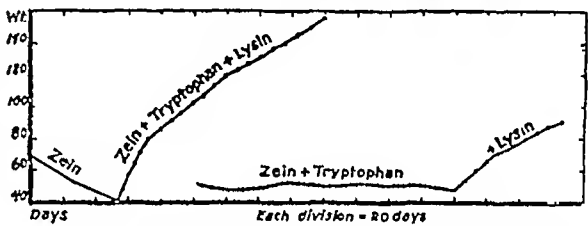


FIG 47 1 Shows the indispensability of lysine for growth (After Mendel)

phaseolin (of navy bean) is deficient in cystine and tryptophane. If these, or other incomplete proteins, are supplemented by adding to the diet the amino-acids in which they are deficient, growth is supported. Also when the amino-acid deficiency is not too great, growth can be promoted by increasing the percentage of the incomplete protein in the diet (see table 61).

An essential or indispensable amino-acid is defined as one which cannot be synthesized in the body in adequate amounts for normal growth and nutrition and must therefore be supplied in the diet. Of the twenty-five amino-acids which have been identified as constituting protein material, ten are indispensable for normal growth. They are as follows:

- | | |
|------------|---------------|
| Threonine | Lysine |
| Valine | Histidine |
| Leucine | Methionine |
| Isoleucine | Tryptophane |
| Arginine | Phenylalanine |

Experiments, in which it was shown that casein

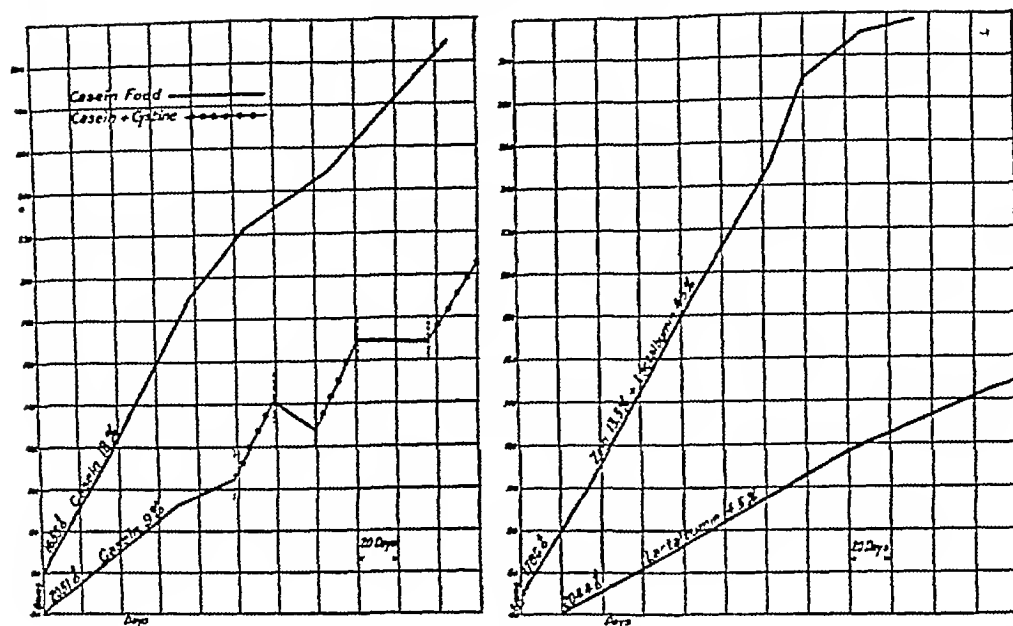


FIG 472 Chart on left shows satisfactory growth of rat when 18 per cent of casein was present in the diet as the sole protein. With a smaller amount of casein—9 per cent—much less rapid growth ensued. That the insufficiency of the smaller amount of casein is essentially due to its relative deficiency in cystine is shown by the marked accelerating influence upon growth brought about by the addition of this amino-acid to the food containing 9 per cent of casein and the prompt retardation of growth which resulted from the withdrawal of cystine from the diet. Chart on right shows the favorable effect upon growth of supplementing a protein (zein), incapable of maintaining animals when it is the sole protein furnished in the diet, with a more "perfect" protein (lactalbumin). The proportion of the lactalbumin used—4.5 per cent—was of itself insufficient for full growth. It evidently furnished the amino-acid groups which were lacking in zein. (After Mendel.)

when constituting the sole protein of the diet would not support growth unless fed in relatively large amounts or in smaller amounts and supplemented by *cystine*, led to the belief that this amino-acid was indispensable. The mistake was evidently the result of confusion with the other sulphur-containing amino-acid, *methionine*. The failure of growth on the low casein diet was due to this protein's inadequate methionine content, but this could be made good by the addition of cystine. In other words, cystine can substitute *in part* for methionine, cystine itself is not essential, for it can be *completely* replaced in the diet by methionine² which can be converted to cystine in the body. Proof of the derivation of cystine from methionine was obtained by Tarver and Schmidt who fed animals methionine containing the radio-

active isotope of sulphur (S^{34}) and later isolated isotopic cystine from their tissues. Cystine is a constituent of plasma protein and unless it or methionine is furnished in adequate amounts in the diet hypoproteinemias result. Cystine is also an essential component of keratin, the protein of hair, feathers, nails, horn, etc. As mentioned on page 630, hemocysteine with choline can, through transmethylation, substitute for methionine. *Arginine* (p 632) can be synthesized in the body but not in adequate amounts for normal growth, it is therefore an essential constituent of the diet of young animals. *Tyrosine* can be formed from phenylalanine. In rats deprived of *valine* grave nutritional defects develop which have been the subject of studies of Epstein and Rose. The animals become very sensitive to touch and show severe muscular incoordination, (e.g., staggering gait and circus movements) together with loss of appetite, emaciation and eventually death.

The amino-acid requirements for building new tissue (growth) are more exacting than those for repair (maintenance), i.e., for the maintenance of

² To make this point clear we may use as illustration the case of a type setter who has lost a number of dies for, say, the letter A. He can continue to set type so long as his supply of A letters lasts, but when this is exhausted his work must cease though there is still a large number of other letters. His type-setting is limited by his supply of A letters.

TABLE 61*
The amino acid content of a number of proteins, grams per cent

AMINO ACID	GELATIN	CASEIN	LACTAL- BUMIN	EGG ALBUMIN	GLIADIN	ZEIN	EDESTIN
Glycine	25.5	0.4	0.4	0.0	0.0	0.0	3.8
Alanine	8.7	1.8	2.4	2.2	2.0	9.8	3.6
Valine	0.0	7.9	3.3	2.5	3.3	1.9	
Leucine-isoleucine	7.1	9.7	14.0	10.7	6.6	25.0	20.9
Aspartic acid	3.4	4.1	9.3	6.2	0.8	1.8	10.2
Glutamic acid	5.8	21.8	12.9	13.3	43.7	31.3	19.2
Hydroxyglutamic acid	0.0	10.5	10.0		2.4	2.5	
Serine	0.4	0.5	1.8		0.1	1.0	0.3
Proline	9.5	8.0	3.8	3.6	13.2	9.0	4.1
Hydroxyproline	14.1	0.2				0.0	2.0
Phenylalanine	1.4	3.9	1.2	5.1	2.3	7.6	3.1
Tyrosine	0.01	6.5	1.9	4.0	3.1	5.9	4.5
Cystine	0.17	0.3	4.0	0.9	2.4	0.8	1.0
Arginine	9.1	5.2	3.0	6.0	3.2	1.8	15.8
Histidine	0.9	2.6	1.5	2.3	2.1	1.2	2.1
Lysine	5.9	7.6	8.4	3.8	0.6	0.0	2.2
Tryptophane	0.0	2.2	2.7	1.3	0.8	0.17	1.5
Total	92.4	94.8	81.9	63.2	91.8	103.4	96.6

* Modified from Mitchell and Hamilton, "The Biochemistry of the Amino Acids," Am Chem Soc Monograph Series No. 48, p. 191

nitrogen equilibrium in the adult and the prevention of a loss of weight. Some amino-acids which are essential for the growing animal can be dispensed with in the adult. Arginine, for example, is not required in the diet of the adult dog, and in the adult rat nitrogen equilibrium can be maintained upon a diet lacking in this amino-acid, but, as just mentioned, it is not synthesized rapidly enough to promote maximal growth. Even of the amino-acids essential for maintenance the level in the diet need not be as high for maintenance as for growth. Adult man can be maintained in nitrogen equilibrium upon nine amino-acids, histidine being dispensable.

The utilization of a protein deficient in one or other amino-acid is limited by that deficiency. For example, if edestin, which is poor in lysine, is the sole protein of the diet, the other amino-acids of which it is composed are utilized only in certain proportions limited by the lysine content. A large part of the remainder is discarded. Similarly the utilization of casein is limited by its cystine plus methionine content. Now if lysine is added to the diet of which edestin is the sole protein, or cystine or methionine is added to the casein diet the other amino-acids in these proteins can be utilized and

built into body tissue.³ Lysine is essential for maintenance as well as for growth, but for the former, relatively small quantities are required. Gliadin (of wheat), which contains less than 1 per cent lysine, is inadequate for growth even though fed in large quantities but is suitable as the sole source of protein in an adult. Zein, on the other hand, which is entirely lacking in lysine and contains little tryptophane is inadequate for either growth or maintenance. Serum albumin is deficient in tryptophane and isoleucine, it will not promote growth in young rats and will scarcely maintain body weight. The proteins of several common foods are given in table 62.

Most of the work relating to the indispensability of the various amino-acids has been carried out upon rats. More recently, Holt and his colleagues have investigated this question in *adult human* subjects. Their results, in general, accord with those of animal experiments. Methionine, as the sole source of sulphur-containing amino-acids, was found to be capable of maintaining nitrogen equi-

³ It follows that a protein of high biological value, i.e., one which possesses an assortment of amino-acids closely resembling that in body protein, i.e., one which when fed alone shows little wastage of nitrogen, will have little supplementing effect.

TABLE 62*

Character of proteins in some common foods

FOOD MATERIALS	CHIEF KINDS OF PROTEIN PRESENT	COMPLETE OR INCOMPLETE
Almonds	Excelsin	Complete
Cheese	Casein	Complete (low in cystine)
	Lactalbumin	Complete
	Glutelin	Complete
	Zein	Incomplete (lacks lysine and tryptophane, low in cystine)
Corn		
Eggs	Ovalbumin	Complete
	Ovovitelin	Complete
Gelatin	Gelatin	Incomplete (lacks tryptophane and tyrosine, only a trace of cystine, high in lysine)
Lean meat	Albumin	Complete
	Myosin	Complete
Milk	Casein	Complete (low in cystine)
	Lactalbumin	Complete
Navy beans	Phaseolin	Incomplete (low in cystine)
Peas	Legumin	Incomplete (low in cystine)
Soy beans	Glycinin	Complete
	Legumelin	Incomplete
	Ghadin	Incomplete (lacks lysine)
Wheat		
	Glutenin	Complete

* From M S Rose, *Foundations of Nutrition*

librium The nitrogen balance could not be maintained upon a diet completely lacking in *lysine*, whereas *histidine* proved to be dispensable Others have found, however, that while nitrogen equilibrium can be maintained in men on a diet lacking in *histidine*, loss of weight occurred The relation of arginine to human nutrition is somewhat ambiguous Holt and his colleagues found that nitrogen equilibrium could be maintained for ten days upon an arginine-deficient diet The seminal fluid of the subjects showed a reduction in the number of spermatozoa When arginine was added to the diet the number of sperm cells was soon restored to normal, but on an ordinary diet several weeks are required for this to occur These authors suggested that the degeneration of spermatozoa, which are very rich in arginine, furnished sufficient amounts of this amino acid to maintain nitrogen equilibrium for a short time, and conclude that in man it is an indispensable constituent of the diet *Tryptophane* was found to be indispensable, the

subjects developing negative nitrogen balances within a few days after having been placed upon a diet lacking in this amino acid, nor could nitrogen equilibrium be maintained when *valine*, *threonine*, *leucine*, *isoleucine* or *phenylalanine* was lacking from the diet *Lysine* and *methionine* were also found to be indispensable, but as mentioned above, adults can be maintained in nitrogen equilibrium in the absence of *histidine*

It seems that, generally speaking, the several amino-acids which have been found indispensable for rats are also essential dietary constituents for other mammals, including man There are, however, certain minor species differences, the growing mouse, for example, can dispense with arginine in the diet, since, apparently, it can synthesize this amino acid in greater amounts than can the rat *Glycine* which is required especially for the manufacture of collagenous tissue and of protoporphyrin (ch 6) can be synthesized by the body and, therefore, is not an essential constituent of the diet in mammals It is indispensable, however, for the growth of chicks, though it can be replaced in this species by creatine The chick, also unlike the rat, has no ability to synthesize arginine The essential or non-essential nature of a particular amino-acid, as well as the species differences, may possibly depend upon bacterial synthesis in the intestine, the bacterial synthesis of amino acids and even of protein is known to occur in the intestinal tract of ruminants It has been found by Martin, for example, that rats fed a diet containing all the essential amino-acids and the known vitamins fail to grow if succinylsulfa-zole, a bacteriostatic drug, is added to the food

The most suitable proteins for growth are those of animal origin and especially those which nature has provided for the nourishment of the growing animal, namely,

lactalbumin (of milk) *ovovitelin* (of hen's egg)
ovalbumin (of hen's egg)

These support growth when given at a level of about 9 or 10 per cent in the diet Next in order of their biological value come,

proteins of meat *glutelin* (maize)
glutenin (wheat) *glycinin* (soy bean)
casein (milk)

These support growth if given in sufficiently high concentration in the diet Casein, for instance, is required to be given at a level of 18 per cent for growth and 10 to 12 per cent for maintenance A

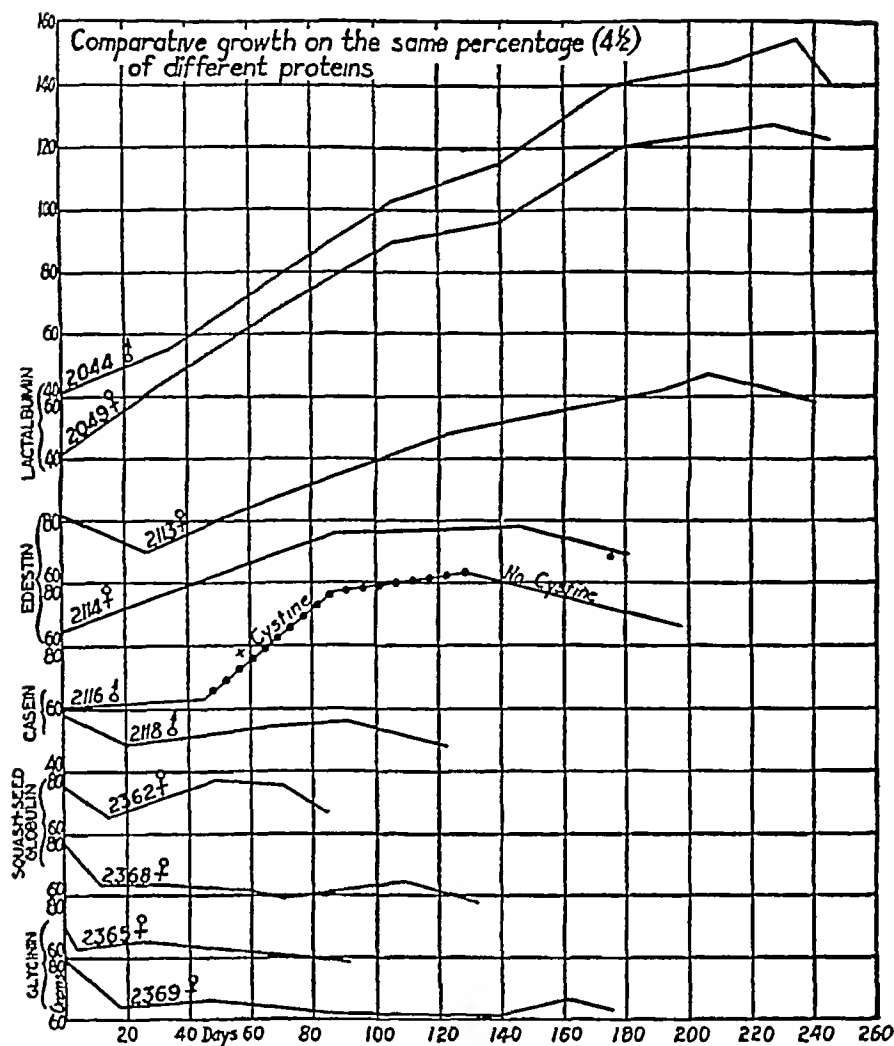


FIG 47.3 Comparison of growth on diets containing approximately the same percentage (4.5 per cent) of different proteins, namely, lactalbumin, edestin, casein, globulin (squash-seed) and glycine (soy bean) (After Lusk.)

comparison of the value for growth of several proteins, when given in the same percentage in the diet, is shown in figure 47.3

The following vegetable proteins are incapable of supporting growth but are suitable for maintenance

- gladin* (wheat or rye)
- legumin* (pea)
- legumelin* (soy bean)
- phaseolin* (kidney bean)
- hordein* (barley)

The following are incomplete proteins and are unsuitable for either growth or maintenance

- zein*
- gelatin*

The biological values of the protein mixtures in the various foods are shown in table 59, p 644, from which it is seen that the values are in de-

scending order as follows eggs, milk, meats, whole wheat and potato, rolled oats, corn, white flour and beans. The nutritional value of ordinary meat varies largely with the cut—a tough fibrous meat, i.e., one with a high proportion of collagen and elastin has a low coefficient of digestibility and a low biological value. The biological value of some samples of very fibrous beef, for instance, may be little higher than that of white flour.

THE SUPPLEMENTARY RELATIONS AMONG PROTEINS

It does not necessarily follow that because a certain protein is inadequate for growth or even for maintenance that it is worthless in nutrition. It is evident from what has been said in the preceding paragraphs that an incomplete protein can be utilized if its shortcomings in one or more of the essential amino-acids are made good by the

addition to the diet of some other protein, rich in the elements which the first one lacked. Zein, for instance, fed in any quantity, is inadequate for either growth or maintenance. Lactalbumin at a level of less than 8 or 9 per cent is incapable of promoting normal growth. Yet normal growth and nutrition are supported by a diet containing 13.5 per cent zein and 4.5 per cent lactalbumin (fig. 47.2). The proportions of tryptophane and cystine and probably also of lysine are higher in lactalbumin than in body protein. Therefore, the excess of these amino-acids in lactalbumin, instead of being wasted, combines with the amino-acids of zein (which otherwise would be discarded) to form tissue protein.⁴

In a similar manner the amino-acids of gelatin are utilized if supplemented by a protein rich in tryptophane and cystine. Even two incomplete proteins, if one contains an abundance of the amino-acids which the other lacks, supplement one another. So, the biological value of two proteins given together may be much greater than the sum of their values when fed

alone. Proteins which lack the same amino-acid (e.g., casein and phaseolin which are deficient in cystine) cannot of course supplement one another. The supplementary relations of proteins is evident in the chief natural foods: these all contain two or more proteins (see table 62). For example, in milk, the deficiency of casein in cystine and methionine is made good by the relatively high percentage of these amino acids in the other milk proteins. Also, though gliadin of wheat will not support growth, wheat itself will. The lysine-poor gliadin is supplemented by the other wheat protein, glutenin, which is rich in lysine. Maize contains, besides zein, the supplementary protein glutelin. These facts, obviously, are of fundamental importance in dietetics, especially in the economy of agricultural feeding. Cheaper foods can be fed and good nutrition promoted if due consideration be given to the supplementary relations of their contained proteins. Gelatin rich in lysine improves the nutritional value of wheat and oats, both relatively poor in this amino-acid. There is also a pronounced supplementary relation between the proteins of milk and those of oats and wheat.

CHAPTER 48

THE METABOLISM OF THE PURINE AND PYRIMIDINE BASES

One class of conjugated proteins—the *nucleo-proteins* (p 623)—are compounds of simple basic proteins, *protamines* and *histones*, with *nucleic acid*. The protein is in salt-like combination with the acid but appears to be present in excess, when a nucleoprotein is hydrolyzed the protein is split off leaving nucleic acid.

Cell nuclei contain nucleoprotein in relatively high concentration, and it was from pus cells obtained from used bandages of surgical wards, and from the heads of salmon sperm, that this conjugated protein was originally isolated by Miescher.

There are two main varieties of nucleic acid. One is known as *thymus-nucleic acid*, the other as *yeast-nucleic acid*. The chemical name of the former is *d-2-desoxyribonucleic acid*,¹ and of the latter, *ribonucleic acid*. They are also referred to as *animal nucleic acid* and *plant nucleic acid*, respectively. But thymus-nucleic acid is not, as was originally thought, confined to animal life, for it has been isolated from the nuclei of plant cells—rye embryos and yeast, plant nucleic acid was found in the cytoplasm. Nucleic acid is constituted of *mononucleotides*. Each of the latter is in turn composed of (1) *phosphoric acid*, (2) a *pentose* and (3) a *nitrogenous group*—a purine, or a pyrimidine base.

A mononucleotide may, therefore, be represented thus

phosphoric acid—pentose—purine, or pyrimidine base

Of the mononucleotides in the nucleic acid molecule some contain a purine base, and others a pyrimidine base.

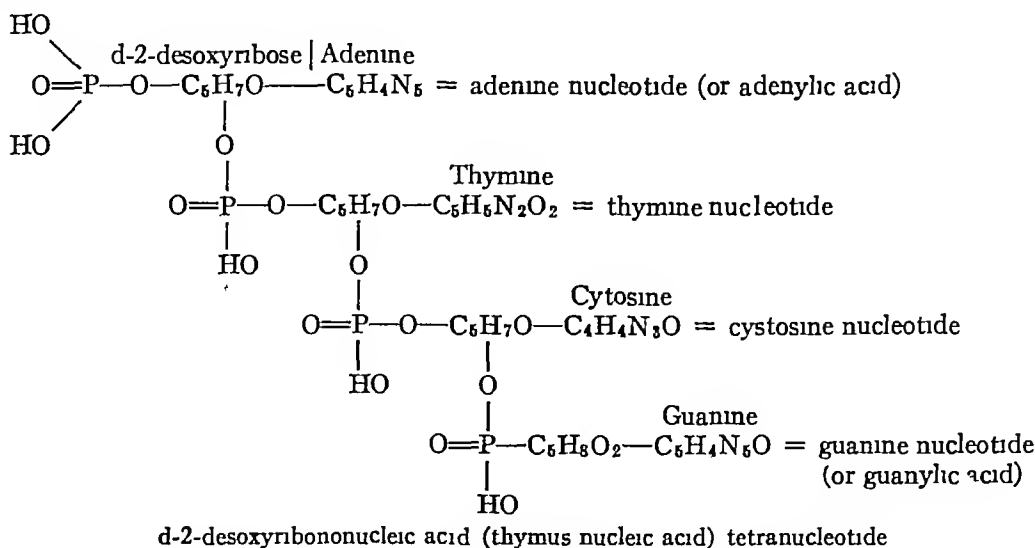
The purines are *adenine* ($C_5H_5N_5$) and *guanine* ($C_5H_5N_5O$).

In thymus-nucleic acid the pyrimidines are *thymine* ($C_5H_5N_2O_2$) and *cytosine* ($C_4H_4N_3O$).

The sugar is *d-2-desoxyribose* ($C_5H_{10}O_4$), a 2-deoxy-pentose.

Plant nucleic acid differs from the latter variety in that its sugar is *d-ribose* ($C_5H_{10}O_5$), and thymine is replaced by another pyrimidine—*uracil* ($C_4H_4N_2O_2$).

The nucleic acid molecule has been generally looked upon as a tetranucleotide, that is, as having four nucleotides linked together. But it now appears that the molecule is much larger (500,000 to 1,000,000 mol wt) consisting of a much greater number of nucleotides which are joined through the phosphoric acid of one nucleotide with the hydroxyl of the pentose of another to form a chain-like structure. It may, therefore, be termed a polynucleotide. The formula of a tetranucleotide, as given by Levene, is shown below.



Upon hydrolysis nucleic acid is split into its constituent nucleotides. The purine nucleotides are termed *adenylic acid* and *guanylic acid*, respectively. The further action of hydrolytic agents splits off phosphoric acid from the nucleotides. The residue (sugar + purine or pyrimidine) is then called a *nucleoside*. The adenine-containing nucleoside is known as *adenosine*, the one containing guanine is called *guanosine*.

THE NUCLEIC ACID DISTRIBUTION IN THE CELLS

The chief nucleic acid of animal cells is desoxyribonucleic acid (abbrev. DNA), smaller amounts of ribonucleic acid (RNA) are also present, largely in the nucleolus. DNA is found mainly in the chromosomes together with small amounts of RNA. The protein with which the nucleic acid is combined is usually a histone, but is a protamine in sperm cells. The cytoplasm of the cells also contains both nucleic acids but RNA predominates. It appears from the ribonuclease test² that the RNA is contained mainly in inclusion bodies (e.g., mitochondria) within the cytoplasm.

There is little doubt that nucleic acids are closely associated with the physiology of cells, and especially of the chromatin material of the nucleus during reproduction. During the different phases of mitosis changes in concentration and distribution of the nucleic acids of the cells occur. In the metaphase the DNA of the chromatin increases while the RNA of the cytoplasm diminishes. In meiosis (sperm cells) the concentration of DNA is doubled during the first division, but not in the second, an increase again occurs in the later spermatid. The nucleic acids are believed to play an essential role in the multiplication of the genes. A number of pathogenic viruses have been shown to be complex nucleoproteins. *Staphylococcus bacteriophage* also is most probably a nucleoprotein.

THE END PRODUCTS OF PURINE AND PYRIMIDINE METABOLISM

Nucleoprotein is split by the digestive enzymes into protein and nucleic acid. Specific tissue en-

² This test depends upon the fact that basophilic granules in the cytoplasm are stained red by pyronine, a basic dye, but fail to do so after incubation with an aqueous solution of ribonuclease. A specific test (Feulgen's) for DNA in cell nuclei consists in treating the tissue with HCl and then immersing it in fuchsin-sulfurous acid reagent (Schill's reagent), in the presence of DNA a purple color appears. By means of these two tests DNA and RNA can be distinguished.

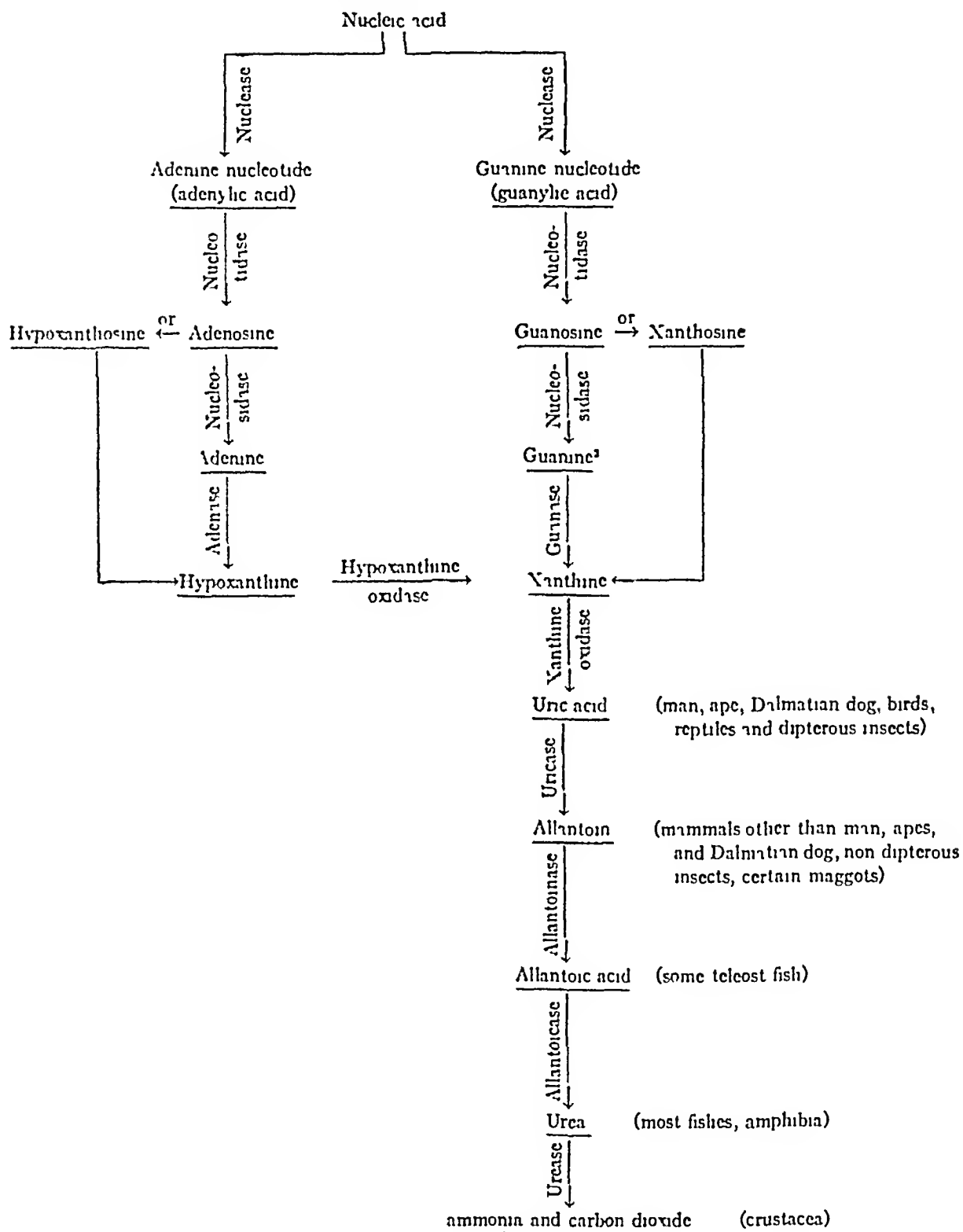
zymes known as *desoxyribonuclease* and *ribonuclease* break the respective nucleic acid molecules into smaller groups of nucleotides—*oligonucleotides* and *tetranucleotides*. These enzymes are present in pancreas and other animal tissues, in yeast, and in certain bacteria (see streptodornase, ch. 12). The further changes which the molecule undergoes in the tissues are not clear, but nucleotidases and nucleosidases which break off phosphoric acid and sugar groups respectively have been described.

URIC ACID (first isolated by Steele in 1776 from urinary calculi). This urinary constituent is the end product of purine metabolism in man (see scheme below), higher apes, birds and reptiles.⁴ Adenine is deaminated by a specific tissue enzyme, *adenase* to form *hypoxanthine*. Hypoxanthine is then oxidized by the enzyme, *hypoxanthine oxidase*, to *xanthine* which in turn is oxidized, by means of *xanthine oxidase*, to uric acid. The other purine, *guanine*, is converted by the deaminative action of *guanase* to xanthine and by further oxidation to *uric acid*.⁵ In mammals other than the primates, and in insects excepting the *Diptera*, from 80 to 98 per cent of the uric acid is oxidized further by an enzyme *uricase*, to *allantoin* and carbon dioxide.⁶ These animals would appear to possess an advantage over man, since allantoin is some 250 times more soluble than is uric acid. The conversion of uric acid to allantoin occurs in the liver (dog). Mann and his colleagues, for example, have shown that after hepatectomy uric acid accumulates in the blood, and when uric acid is injected, from 70 to 100 per cent can be recovered from the urine unchanged. This is in marked contrast to the behavior of the normal dog which excretes 98 to 100 per cent as allantoin. Extracts of dogs' liver are rich in uricase.

⁴ In birds and reptiles uric acid is the end product of protein metabolism as well as of purine metabolism. Arginase is absent from the liver of these forms. According to Benedict from 60 to 70 per cent of the urinary nitrogen of these species is in the form of uric acid. Thus, the end product of nitrogen metabolism differs according to the animal class. Those animals, birds and reptiles in which the main end product is uric acid, are called *uricotelic* (Gk. *telos*, end), whereas, mammals which excrete nitrogen mainly as urea, are termed *ureotelic*.

⁵ The production of uric acid may follow another course, namely, conversion of the nucleoside adenosine or guanosine, by deamination, to hypoxanthosine or xanthosine, which is then split into its pentose and purine groups, hypoxanthine or xanthine, respectively (see plan on p. 653).

⁶ The spotted coach-dog (Dalmatian) oddly enough is an exception like man and unlike other canine breeds it excretes uric acid.



In some classes of ureotelic animals (most fishes and amphibia) further degradation of uric acid

³ Guanine is the end product of purine metabolism in the spider

occurs, allantoin being converted to allantoic acid (catalysed by allantoicase), and then to urea. In crustacea, the latter is finally broken down to ammonia and carbon dioxide.

The stages through which uric acid or allantoin is derived from nucleic acid are shown in the scheme on page 653

The chemical changes involved in the conversion of guanine and adenine to uric acid or allantoin are indicated in the formulae (page 655)

Little is known concerning the fate of the pyrimidines in man. In animals they are for the most part completely oxidized, the nitrogen being excreted mainly as urea. The urine contains pyrimidines in insignificant quantities.

THE SYNTHESIS OF PURINES IN THE BODY

There is no doubt that purine synthesis occurs in the young mammal, for, when upon a diet comprised exclusively of milk, which is almost purine-free, it excretes uric acid or allantoin, and at the same time manufactures nucleic acid for the construction of cell nuclei. That synthesis also occurs in the adult was shown by Benedict, who kept a Dalmatian dog upon a purine-free diet for nearly a year, during which time the animal excreted 100 grams of uric acid. This had not apparently been derived from body tissue since the animal maintained a constant weight. In more recent studies with isotopically labelled nitrogen (N^{15}), purine synthesis has been proved conclusively, after the administration of amino acids or ammonium salts containing N^{15} the isotope has been found in the nucleic acids of the animal's tissues. Glycine has been demonstrated in the same way, to participate in the formation of the purine ring.

Histidine and arginine are thought to be important precursors of purines. Rats fed upon diets deprived of both these amino acids cease to excrete allantoin.

THE PRODUCTION, DESTRUCTION AND EXCRETION OF URIC ACID

In man, uric acid is formed in part from purines taken in the food (*exogenous uric acid*) and in part from body purines (*endogenous uric acid*). The latter are derived from the break-down of nuclear material as well as from free mononucleotides, e.g., adenylic, inosinic and guanylic acids, found in muscle and glandular structures. Muscle and more especially glandular tissues, e.g., thymus, liver, kidney, pancreas, testes, etc., and leguminous vegetables are food rich in purines.

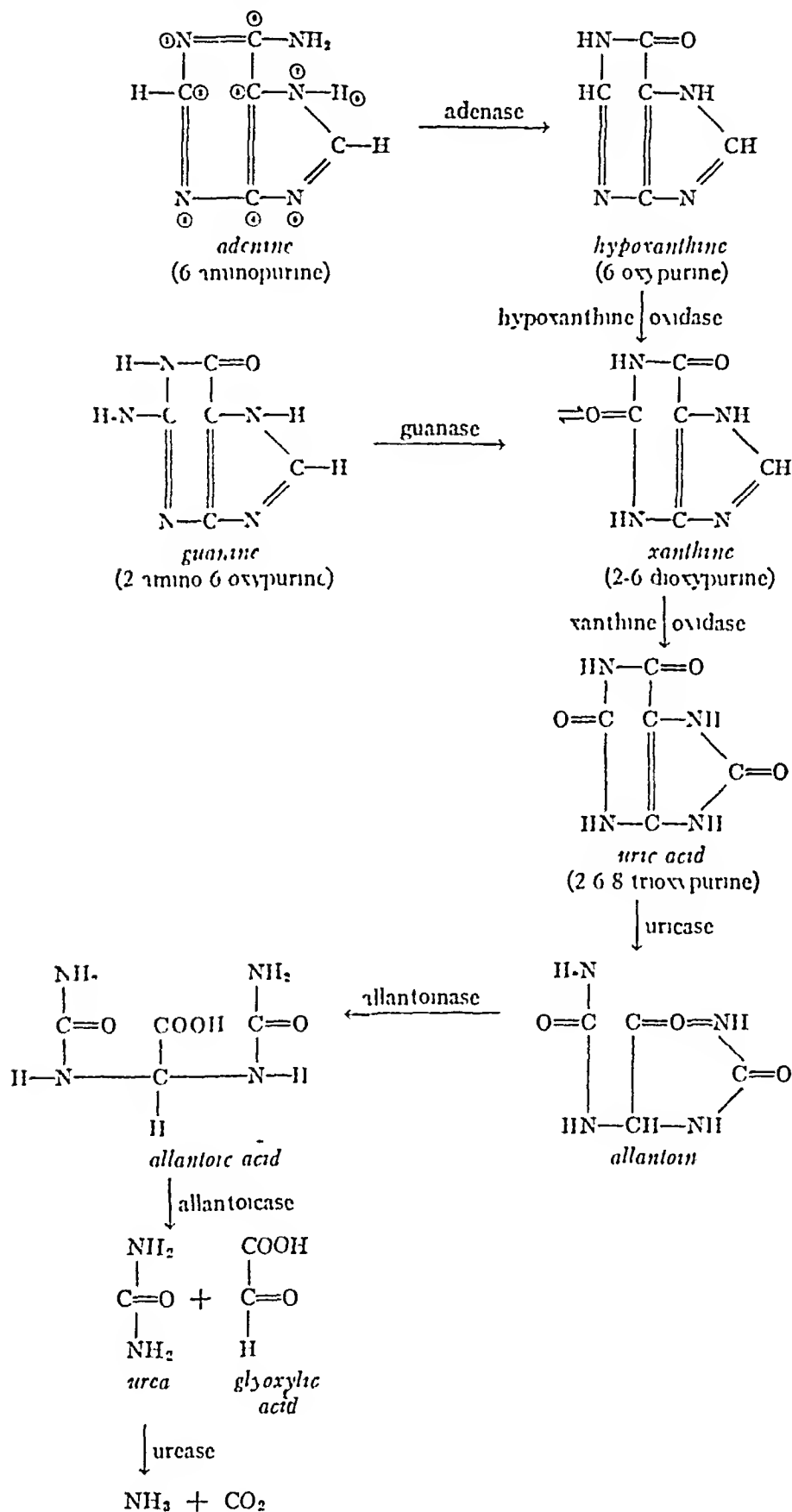
Uric acid is formed from ammonia in the liver of the bird but little is known of the intermediate stages of the synthesis. As already mentioned (footnote, p. 653) uric acid is the end product of both protein and purine metabolism in birds and most reptiles, when ammonium salts containing N^{15} are fed to birds, uric acid labelled with the

isotope is excreted. According to Krebs and his associates hypoxanthine is first produced. Just as in the dog no urea is formed after hepatectomy, so in birds, no uric acid is produced after removal of the liver, ammonia accumulates in the blood. The site of uric acid production in man is unknown, but the liver, according to Jones, is the only tissue which contains xanthine oxidase, which suggests that uric acid production is, in part at least, a function of the human liver. The results of the investigations of Folin and associates indicate that from 30 to 70 per cent of the uric acid produced by the human subject is destroyed, the remainder appearing in the urine. The liver is probably the site of uric acid destruction in man. We have already seen that uric acid accumulates in the blood of the hepatectomized dog, which indicates that uric acid production occurs in extra-hepatic tissues and that the uric acid/allantoin conversion is a function of the liver. Little information is available concerning the process whereby uric acid is destroyed in the human body, its end products are unknown, allantoin is not one of them however, since human urine contains insignificant amounts of this substance, and what little is present is simply that taken preformed in the food. Uric acid is excreted in the urine as the urates of sodium, potassium and ammonium, and in the free state. After urine has been voided a crystalline deposit of urates and free uric acid appears. The average daily output of uric acid, on an ordinary diet, is from 0.5 to 1 gram. Of this the endogenous uric acid amounts to from 0.3 to 0.4 gram. Muscular tissue most probably furnishes the precursors of endogenous uric acid, exercise causing a rise in the uric acid of the blood, and increased excretion in the urine.

Protein (purine-free, e.g., milk, eggs) and carbohydrate foods accelerate the excretion of uric acid, and lower its level in the blood.⁷ Food fat has the reverse effect—reduced excretion and a raised level in the blood.

Small quantities (15–45 mg. daily) of purines are excreted in the urine. These include adenine (but not guanine), hypoxanthine and xanthine, as well as the methyl purines contained in beverages. The latter are caffeine (1,3,7-trimethyl xanthine) of coffee, *theo-*

⁷ The possibility has been suggested in the past that a proportion of the uric acid formed in the human body is derived, as the great part of it is in birds, from the catabolism of protein. There is no reason to believe, however, that a conversion of this nature occurs to any significant extent.



phylline (1,3-di-methyl-xanthine) of tea and *theobromine* (3,7-di-methyl-xanthine) of cocoa. The methyl purines undergo partial demethylation in the body and are excreted as mono- and di-methyl purines. It appears that theobromine is excreted entirely in this form and is not converted to uric acid. Caffeine and theophylline on the other hand are converted in considerable amounts to uric acid and excreted as such. This is an important point to bear in mind when considering dietary restrictions for gouty subjects.

GOUT

The uric acid content of normal blood (both free and bound with protein) averages about 3 mg (2.5 to 5.0 mg) per 100 cc. In gout the concentration of uric acid in blood is raised and may be as high as 10 mg per cent while its excretion in the urine is reduced. The excretion of uric acid after a meal rich in nucleoprotein, as well as the endogenous uric acid excretion, that is, the excretion on a protein-free diet, are much less in the gouty than in the normal subject. The diminished excretion of uric acid is, as a rule, particularly well marked just prior to an acute attack of gout. There may be no rise, however, in the blood uric acid at this time. A characteristic feature of gout is the formation of deposits of the sodium salt of uric acid (sodium urate) in the form of crystals in the cartilaginous tissues. The deposits are found most commonly about the joints of the great toe and fingers. They also occur in the helix of the ear or the tarsal plates of the eyelids, in which situations they are known as *tophi*. During the gouty attack the joint becomes acutely inflamed and tender. The adrenocorticotrophic hormone, or cortisone, increases uric acid excretion and has been employed in the treatment of the disease.

There is no reason to believe that the *hyperuricemia* of gout is the result of an increased production of uric acid in the body, nor is it due, apparently, to its diminished destruction. It is quite evidently due to reduced excretion. Nevertheless, the uric acid retention is not due to renal failure in the ordinary sense of the term, and though gout and kidney disease are not infrequently associated, the high concentration of uric acid in the blood of gouty subjects is the forerunner of the kidney damage, rather than that the latter is the primary cause of uric acid retention. Furthermore, a level of uric acid may be present as in nephritis and other conditions without any sign of gout. These facts make it difficult to find a satisfactory explanation for the high uric acid level in the blood in gout. Minkowski suggested that normally uric acid existed in blood in combination with a nucleoside which was

decomposed by the kidney cells, the uric acid being excreted. In gout, it was supposed, this specific renal function was impaired. Benedict and his colleagues have obtained evidence in favor of such a view. They claim to have demonstrated the presence of two forms of uric acid in the blood of the ox and other animals, and in man. Freshly drawn ox blood contains 0.5 mg per cent of uric acid. Ten times this amount was obtained when the blood had been allowed to stand, or had been boiled with hydrochloric acid. This suggested that the uric acid was present largely in some stable, combined form which was decomposed by the action of an enzyme or by treatment with acid. The combined form is present exclusively in the erythrocytes. It is made up of a molecule each of uric acid and pentose—d ribose.

The deposition of urates in the tissues is not simply the result of the saturation of the blood with uric acid. The concentration of uric acid in the general circulation never reaches the limit of its solubility in blood. In other conditions, e.g., nephritis and leukemia, associated with a high blood uric acid, deposits do not occur. Some local factor apparently being necessary. The process is possibly in the nature of vicarious excretion of uric acid into certain tissues (Benedict) wherein the uric acid, upon reaching the saturation point, is deposited as crystals of sodium urate.

It is the usual practice to treat gout by rigid dietary prohibitions, e.g., reduction in the intake of purine-containing foods and fat, replacing them by carbohydrate and such purine-free proteins as in milk and eggs which are believed to increase uric acid excretion (p. 654). Salicylates, colchicum, aspirin and cinchophen (atophan) which are used in the treatment of gout have been shown to increase uric acid excretion, and to lower the level of blood uric acid. Alcohol appears to exert little effect upon the latter though it is usually banned from the diet. Coffee and tea, since their methyl purines are partly excreted as uric acid, are restricted.

Bauer and Klemperer have thrown doubt upon the benefit to be derived from an extremely austere diet in gout. They found that the curtailment of the purine intake below that in an ordinary diet exerted no influence upon the course of the disease. Gouty attacks were not induced by a high fat diet, nor were they prevented by a diet free from fats. The only articles of diet which they consider desirable to restrict are those very rich in purines, e.g., sweetbreads, liver, kidney, anchovies, sardines, etc.

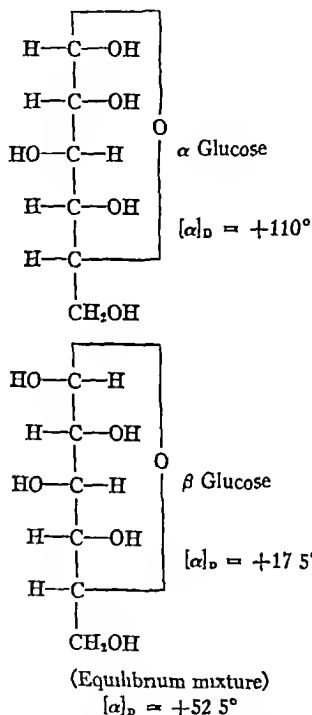
Other conditions associated with a high level of blood uric acid are

(a) *Leukemia and polycythemia*. The elevation

of the blood uric acid in these conditions is not associated with symptoms of gout. The high level of blood uric acid is evidently due to increased production and not to failure in excretion. In leukemia, for example, the amount passed daily in the urine may be as much as 12 grams. The increased uric acid production in this condition is probably due to the disintegration of body tissue generally rather than, as has been supposed, to the destruction of white cells. (b) *Pneumonia* (c) *Nephritis*. In renal failure the uric acid is one of the first of the nitrogenous substances of the blood to show a rise. (d) *Lead poisoning* (e) *Toxemia* of pregnancy (f) In certain other conditions associated with a high non-protein nitrogen of the blood.

CARBOHYDRATE METABOLISM

Without doubt the carbohydrate substance most important in metabolism is glucose. All organisms appear to be able to utilize glucose. Glucose, sometimes called dextrose because its solution rotates the plane of polarized light in a dextro-rotatory direction and often written d-glucose, dissolves in water to give an equilibrium mixture of α and β -glucose, approximately two thirds of the glucose being of the α form. The formulae of these two isomers are written



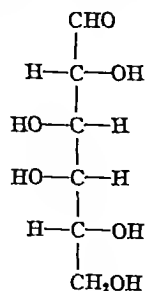
of glucose are very reactive, that is, unstable. They have not been isolated. It may be pointed out, as an example of their reactivity, that they reduce Fehling's solution in the cold. Because of this marked reactivity it has been postulated that such compounds must be the first intermediaries in glucose metabolism. But there is no conclusive evidence for such a change, and all theories of glucose metabolism involving the formation of a "reactive" form of glucose have very little factual basis. Glycogen, or animal starch, is a polysaccharide composed of glucose units and so is written $(\text{C}_6\text{H}_{10}\text{O}_5)_x$ [according to Haworth, $x = 12$. In liver glycogen formed from galactose x may equal 18 (Bell)]. Glycogen is widely distributed in the animal body, but the bulk occurs in the muscles and the liver. Glycogen may be isolated from tissue by hot water extraction, or more readily and completely by hot alkali. It is worthy of note that whereas polysaccharides are very stable in alkaline solution, monosaccharides (glucose, fructose, galactose, etc.) are unstable. The reverse behavior is exhibited in acid solution. The glycogens, as Cori has shown, are a series of branched polysaccharides differing mainly in the degree of branching and the length of the so-called inner and outer chains. Step-wise degradation by the action of phosphorylase and amylo-1,6 glucosidase and synthesis by phosphorylase and the branching enzyme of liver and muscle has revealed the structure.

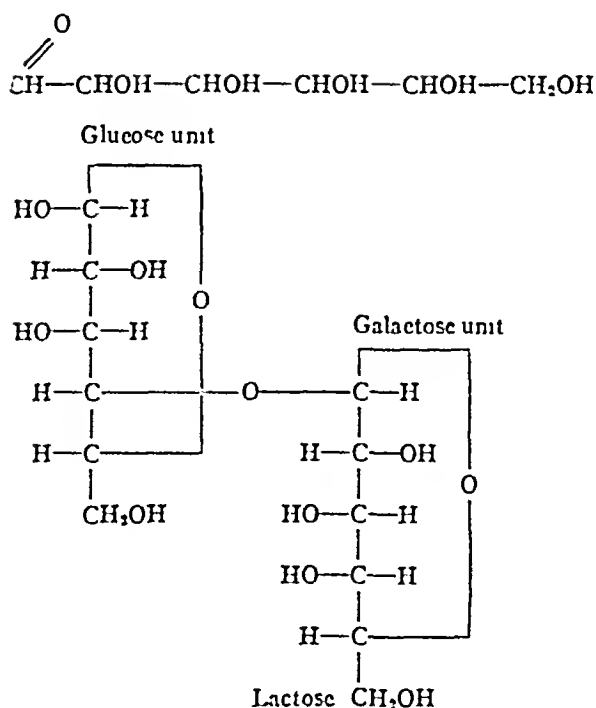
Glucose occurs in the body in combination with other substances, such as salts of glucose phosphoric acids, and in the lactating animal in combination with galactose as lactose, which is α - and β glucose β galactoside.

It will be noted that a cyclic structure has been accorded the glucose molecule.¹ In this case the ring is composed of five carbon atoms and one oxygen atom, sometimes referred to as an aprotene oxide ring. α and β isomers of glucose may also possess a ring composed of four carbon atoms and one oxygen atom. This is the so-called γ -glucose. Such a ring is a butylene oxide ring. These forms

¹ To account for the characteristic reducing properties of glucose it is presumed that in solution there exists a very small amount of the open chain aldehydic form in equilibrium with the cyclic forms.

Galactose (open chain form)





The slight difference in chemical configuration between glucose and galactose should be noted. Carbohydrates occur as substituent groupings in certain proteins. Such proteins may be extracted from tissues paying due regard to their solubility properties. From such extracted proteins there has been isolated glucose (from glycoproteins) glucosamine, mannose (from certain blood proteins) and various pentoses, that is 5 carbon sugars possessing the general formula ($\text{C}_5\text{H}_{10}\text{O}_5$) in nucleic acid which is combined with certain proteins to give the so-called nucleoproteins (ch 49). It may be pointed out that pentoses also occur in combined form as nucleotides, and that very rarely cases of human pentosuria are encountered.

Carbohydrates also occur linked with fatty substances. Thus the cerebroside contains within their molecule galactose units. The galactose is present in the normal or amylenic cyclic ring form. Cerebrosides may be extracted with the usual fat solvents. The galactose may be split off from such extracted substances by acid hydrolysis. Glucose may also be present in cerebroside (p 689).

METHODS OF ESTIMATION

Glucose. The simple sugars have three properties which permit their identification and estimation: (1) their ability to reduce the salts of heavy metals in alkaline solution, (2) their optical activity, (3) fermentation by yeast and by various microorganisms. The methods depending on the reduction of the salts of heavy metals

have been of great value in physiological work but since other compounds present in the body also possess this property it is necessary to prove that the reduction is due to glucose. This is accomplished by utilizing the third property mentioned above, i.e., by fermenting the glucose with yeast or with certain microorganisms. The optical activity of the carbohydrates is an invaluable property when they are present in solutions uncontaminated by various other optically active substances which are present in physiological fluids.

Glycogen. Glycogen is very resistant to alkaline hydrolysis, but, as Claude Bernard discovered, readily yields reducing sugar on acid hydrolysis. Pflüger boiled glycogen for several weeks in strong alkali without destroying it. Advantage is taken of this property to free glycogen of contaminating substances. After alkaline hydrolysis of the tissue the glycogen is precipitated from solution by alcohol. This material is then broken down by acid hydrolysis to glucose, which is stable in acid solution. The sugar is then estimated as described above and the amount of glycogen calculated from the result. The histochemists using the most effective "vital" staining procedures will probably soon be able to tell us the approximate amount of particulate glycogen in the living cell and perhaps also of the soluble form which is, of course, one step nearer the metabolic pathway.

Lactic acid. In muscle, glycogen is broken down to lactic acid during muscular exercise and in the recovery from exercise part of it is resynthesized to glycogen. The lactic acid is estimated by oxidation to acetaldehyde and the latter is determined by distilling it into sodium bisulphite with which it forms a double-compound. Residual unbound bisulphite is titrated with iodine.

APPROXIMATE DISTRIBUTION OF CARBOHYDRATES

	Liver per cent	Muscle per cent	Blood per cent
Glucose	0.06-0.15	0.02-0.04	0.08-0.11
Glycogen	0.2-10.0	0.2-1.8	trace
Lactic acid	0.01	0.01	0.01

Human liver is about 3 per cent, muscle 50 per cent and blood 7 per cent of body weight (Brody and Kibler). Glucose and lactic acid are found in the other soft tissues and some of these contain small amounts of glycogen. The glycogen content of kidney and of heart muscle has been extensively studied under various physiological and pathological conditions.

ABSORPTION OF SUGAR

The monosaccharides formed during digestion are rapidly absorbed from the small intestine. In the rat, galactose disappears from the intestine most rapidly, glucose a little more slowly and fructose much more so. There is, therefore, a selective

action of the intestinal cells in the absorption of sugars (Con) Evidence is accumulating that the absorption of sugar may proceed against a concentration gradient, i.e. when the concentration is lower in the lumen of the small intestine than in the blood. There is an increase in the amount of esterified phosphate in the intestinal mucosa during the absorption of both glucose and fructose but this phosphorylation may be related to metabolism rather than to absorption. It has, however, been suggested by Lundsgaard that the relatively high level of esterified phosphate in the intestinal mucosa during the absorption of fructose may be due to slow dephosphorylation and thus account for the well known delay in the rate of absorption of this sugar.

The slow absorption of glucose from the small intestine after hypophysectomy or adrenalectomy may be a part of the picture of inanition rather than a result of the loss of a specific adrenal phosphorylating factor, as has been suggested. The fact that the administration of sodium salts restores the rate of glucose absorption in adrenalectomized animals (Deuel) is in favor of the non-specific effect.

Glucose is not absorbed in appreciable amounts from the stomach until high concentrations are reached, and only slowly from the large bowel.

THE FATE OF INGESTED OR INJECTED GLUCOSE

There is good evidence that no other monosaccharide is as effective in the liverless animal as glucose. This finding suggests, and there is indeed good evidence, that the other monosaccharides are normally changed to glucose in the liver. Fructose is a better glycogen former than glucose but galactose is much inferior in this respect. When glucose is absorbed or is injected some of it can be accounted for by the increase in liver glycogen, some is converted to muscle glycogen. The concentration of glucose in the soft tissues is temporarily raised. The use of isotopes has revealed a rapid and extensive conversion of glucose to fat (Stetten) and a more delayed increase in oxidation (Wick and Drury) (see page 671). These processes, but not the precise time relationship, had been well established previously.

THE FORMATION OF GLUCOSE IN THE BODY

Glucose is apparently formed in large amounts only in the liver. When this organ is removed (Mann and Magath) hypoglycemia soon appears

and sugar must be provided if the animal is to survive more than a few hours. Sugar is made from protein in the liver and evidence obtained on phloridized animals indicates that some of the amino acids, glycine, alanine, cystine, aspartic and glutamic acids may yield glucose in the theoretical amounts (Lusk). It is well established that plants form sugar from fat. There are a certain number of reports of respiratory quotients below 0.7 which disturb those who believe in the unitarian significance of the quotient, but this ratio should not be used as a weapon against itself. It has been established that an appreciable amount of sugar is formed in the kidney (Bergman and Drury, Reincke), but this appears to be a small fraction of that supplied by the liver.

The Common Metabolic Path

The recent discovery of the nature of the active "2-carbon" fragment, which has long been postulated as being an important intermediate in the metabolism of fatty acids, indicates pathways by which the conversion of carbohydrate to fat (or the reverse change) occurs. The labile acetyl groups found in acetyl-coenzyme A which arise in the oxidation of pyruvate or from the oxidation of fatty acids, appears to be identical. It is not surprising, therefore, in view of this common pathway in the metabolism of carbohydrate and fat to note that the old controversy as to whether or not conversion of fatty acid into glycogen can occur has been settled in the affirmative by the use of isotopes. The small amount of isotopes incorporated into glycogen agree, however, with the view that fat is not quantitatively an important precursor of glycogen or glucose. Since certain amino acids form sugar it is becoming increasingly clear that the various metabolic materials share a common pathway and that acetyl-coenzyme A provides a link which joins them (see Fat Metabolism).

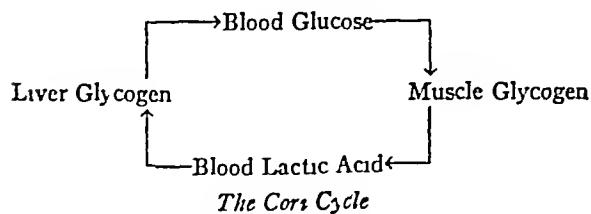
FORMATION AND BREAKDOWN OF GLYCOGEN AND THE OXIDATION OF SUGAR

Great advances in our knowledge of the most fundamental aspects of carbohydrate metabolism have been made in recent years. Most of the steps in the formation and breakdown of glycogen and the oxidation of sugar have been revealed by the work of Harden, Young, Embden, Meyerhof, Warburg, Cori, Lohmann, Peters and numerous other investigators. Many of the individual steps are not observed in the intact cell since the intermedi-

ate products which have been identified "in vitro" do not accumulate. There are, for example, at least twelve enzymatic reactions involved in the anaerobic conversions of glycogen to lactic acid. Practically all of these have been shown to be reversible but since the final product of one reaction is immediately removed by the following step the process proceeds in one direction. The phosphorylation of glycogen and of glucose has been proved to be the introduction to a long series of changes by which these products are transformed through various phosphate esters to pyruvic or lactic acid. Glycogen after phosphorylation breaks down to glucose-1-phosphate. This (Cori ester) is converted to glucose-6-phosphate which is also the first step in the oxidation of glucose.

The work of Cori and Cori in which the enzyme phosphorylase, which catalyzes the reversible reaction $\text{glycogen (or starch)} + \text{inorganic phosphate} \rightleftharpoons \text{glucose-1-phosphate}$, was isolated and purified, is of great interest. This enzyme is widely distributed in animal tissues and is responsible for the first stage in the breakdown of glycogen. It also affects the synthesis of a polysaccharide indistinguishable from glycogen. Some glycogen must be present to "prime" this latter reaction and adenylic acid is an essential constituent of the mixture for activity in either direction. It is now well established that this transfer of phosphate groups is of paramount importance in the mechanism by which cells derive energy from the breakdown of food materials. Muscle glycogen breaks down to CO_2 and H_2O under physiological conditions. The steps from glycogen to pyruvic acid can occur anaerobically but from pyruvic acid on, an adequate oxygen supply is necessary. When there is a relative lack of oxygen lactic acid is formed. Pyruvic and lactic acids are interconvertible in the body. This lactic acid may diffuse out into the blood stream and it is then converted in the liver to glycogen.

This part of the cycle is apparently not brought into play under ordinary circumstances but is involved in muscular exercise or under conditions of anoxia.



Glycogen-containing tissues, with the exception

of the liver, exhibit the same pattern of breakdown of this polysaccharide as has been described in muscle. In the liver, glycogen does not normally break down to CO_2 and H_2O or to lactic acid but to dextrose. This is probably due to the fact that the liver contains a very active phosphatase which converts the glucose-1-phosphate and the glucose-6-phosphate to glucose and thus removes these esters from the medium.

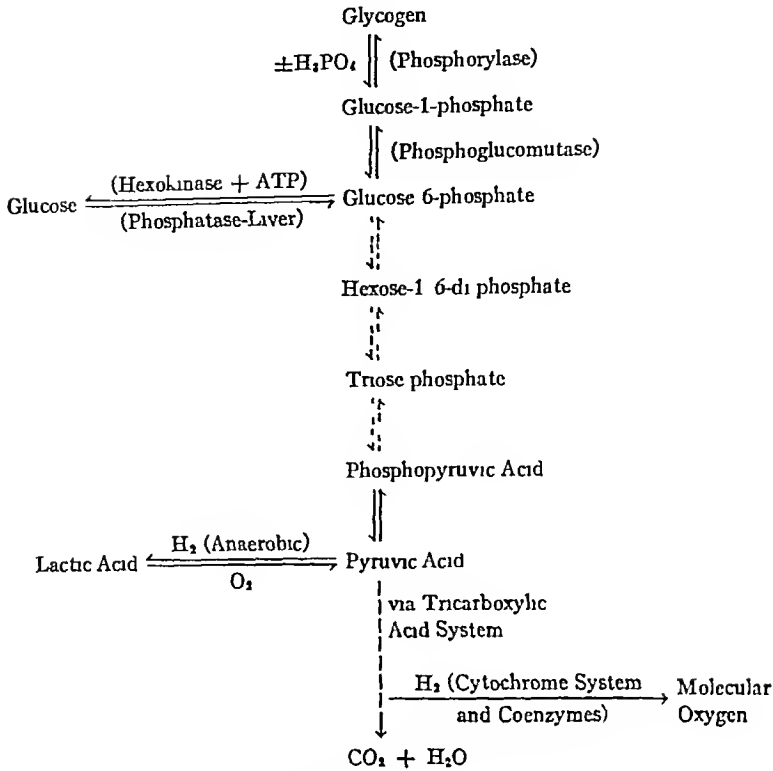
Under certain conditions, *in vitro*, glyceric aldehyde phosphate and dihydroxyacetone phosphate can be isolated from the breakdown products of muscle glycogen. In the diabetic dog these substances, and also pyruvic acid, form "extra sugar", but so does methyl glyoxal which is not a part of the accepted schemes.

THE REGULATION OF BLOOD SUGAR

The blood sugar level represents the resultant of oxidation, storage and excretion on the one hand and formation and absorption on the other. This may be represented graphically.

The relative constancy of the blood sugar which is an equilibrium mixture of α and β glucose in the normal fasting animal is made more remarkable by the fact that the production, storage, and utilization of sugar are affected by a great number of chemical and nervous factors.

When sugar is not being absorbed from the intestinal tract it must be made in the liver or kidney from non-carbohydrate sources or liberated by the breakdown of glycogen. The liver glycogen is the emergency supply which is available while the process of gluconeogenesis is gathering speed. In the average man approximately 100 gm of glycogen are present in the liver and this would only supply the demands for sugar for some five hours if the gluconeogenesis were to cease. The rate of glycogen breakdown is affected by adrenaline which is liberated in emergencies (see also glucagon) and the rate of gluconeogenesis by the internal secretions of the anterior pituitary gland, the adrenal cortex, the thyroid and the pancreas, as will be discussed later. An important part of this homeostatic mechanism for the regulation of blood sugar appears to be the level of blood sugar itself (Soskin). Under experimental conditions in which variation in the rate of secretion of insulin is impossible, i.e. in the depancreatized dog given insulin at a constant rate, indirect evidence indicates that the administration of dextrose decreases the rate of output of sugar by the liver. In the

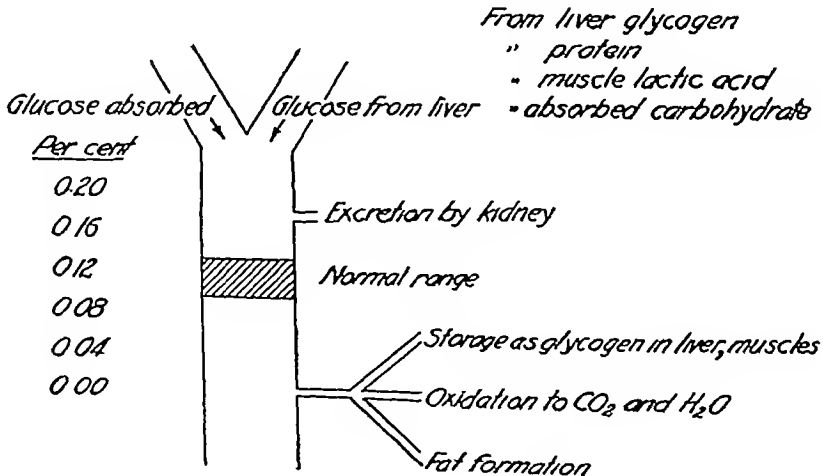


(Other intermediate steps in this process shown in ch 52)

intact animal the pancreatic mechanism is also involved. The insulin content of the blood probably determines at what level of blood sugar gluconeogenesis in the liver will be inhibited. In the absence of insulin only extremely high levels of blood sugar exert a protein sparing action.

HYPERGLYCEMIA

An increase in the blood sugar beyond the normal range constitutes a hyperglycemia. This may be a protective phenomenon. With the exception of insulin, a rise in blood sugar provides the greatest



single stimulus for the formation of glycogen and the utilization of glucose

Prolonged hyperglycemia may be present when either or both of the following processes are operating (1) diminished utilization of glucose, (2) overproduction and discharge of sugar from the liver. Disturbances in endocrine function are the main causes and the mechanism of these changes will be discussed later.

Transient hyperglycemia may be due to either physiological or pathological processes. Alimentary hyperglycemia is a physiological process and its height depends on the amount and nature of the carbohydrate in the meal. Adrenaline hyperglycemia is a part of the physiological response to an emergency. Nerve impulses acting on the liver, changes in hydrogen-ion concentration of liver cells due to asphyxia or other causes, may produce a transient rise in blood sugar. Toxic products of infection acting on the liver to increase gluconeogenesis or on tissues in general to diminish utilization of sugar may produce a hyperglycemia of short or long duration. The physiological role of glucagon is being actively studied.

Glucosuria

The occurrence of glucose in the urine is the result of hyperglycemia except when the renal threshold to glucose is lowered (phloridzin diabetes in animals, renal diabetes in man). The following list of the methods of causing experimental glucosuria is slightly modified from Ingle.

1 Pancreatic (insulin) insufficiency

Pancreatectomy—Von Mering and Minkowski

Alloxan—Dunn, Sheehan and McLetchie
Subtotal pancreatectomy and high caloric intake—Allen

Withdrawal of insulin following its prolonged injection in the force-fed normal rat—Ingle et al

2 Hormonal

Crude anterior pituitary extract or purified growth hormone—Houssay and Biasotti, Evans, Young, Campbell, Anderson

Adrenal C-11-oxygenated steroids—Long, Katzin and Fry, Ingle

Estrogens—Dolin, Joseph and Gaunt, Ingle

Thyroid—Houssay

3 Dietary

Starvation and subsequent feeding—Lehmann

Overfeeding—Hofmeister

Sudden shift from high fat to high carbohydrate diet

4 Glycogenolysis

Piqure—Bernard

Emotion—Cannon, Shohl and Wright

Epinephrine, glucagon, drugs, toxins, trauma

5 Renal

Phloridzin—Von Mering

PANCREATIC DIABETES

Removal of the pancreas

When the pancreas is completely removed from a dog or cat, a characteristic syndrome rapidly develops (Mering and Minkowski, 1889). When a diet including the known essentials is provided, the animals may live indefinitely if adequate amounts of insulin are administered. The animals recover rapidly from the operation and appear normal, but the diabetic state quickly supervenes when insulin is discontinued. The sugar content of the blood begins to rise within a very short time, depending on the size of the last dose of insulin, and increases from the normal level of 0.08–0.11 per cent to 0.20–0.40 or higher within twenty-four hours. The urine gives a positive Benedict's qualitative test for sugar when the blood sugar rises above approximately 0.16 per cent. This point, the so-called renal threshold, is the level of blood sugar above which large amounts of sugar are excreted in the urine. In some animals the "threshold" rises when the diabetic state is allowed to persist. (There is some glucose in normal urine and small amounts of other sugars.) The excretion of nitrogen is increased and this may be taken to indicate protein breakdown or decreased protein formation. The ratio of glucose to nitrogen excreted in phloridzin poisoning is 3.6:1 but this ratio is not observed in depancreatized dogs and indeed that obtained is so variable that little significance can be attached to it. The ratio is low in animals on a high protein diet and there is evidence that the addition of fat may increase it somewhat. It is established that glycerol may be converted to sugar. In the fasting diabetic the blood sugar and sugar excretion are maintained at high levels. The sugar is apparently formed from body protein in the liver since the blood sugar of the diabetic falls rapidly after hepatectomy. The amino acid contents of the blood and urine are increased. The loss of protein contributes to the decrease in body weight.

The disturbed metabolism of fat in the depancreatized animal is indicated by the accumulation of the ketone bodies in the blood and by the excretion of excessive amounts in the urine. The ketosis in a fat dog is greater than in a lean, but this species is characterized by its efficiency in metabolizing fats, without ketosis. The loss of body fat is rapid but the ketosis may be so severe, even in this species, that the animal dies in coma before the fat reserves are depleted.

β hydroxy-butyric and aceto-acetic acids and others derived from tissue breakdown appropriate base and thus when the available reserve of base is depleted may produce an acidosis. In acidosis the respiratory center is stimulated and "air hunger" and coma are produced. The mechanism of coma production in diabetes is not completely known. Some observers believe that acetoacetic acid is particularly toxic. This acid is oxidized in the bladder and lungs to form acetone, which is excreted in the urine and expired air.

The neutral fat content of the blood increases, due probably to the increased rate of mobilization of depot fat and there is also a rise in cholesterol esters and the phospholipid.

While there is proof that the depancreatized dog can still burn sugar—all the criteria of sugar combustion in the normal have been satisfied—it is not permissible to conclude that there is not interference with this process. The rapid sugar utilization in the hepatectomized diabetic animal and the production of a diabetic condition by the administration of cortisone, which apparently acts, in part, by stimulating the liver to produce more glucose, tend, however, to emphasize over-production rather than under-utilization. On the other hand the anterior pituitary and the adrenal cortex contain substances which inhibit the utilization of glucose. Many of the studies on this controversial problem have been made on resting or highly abnormal preparations. The evidence which is now accumulating from studies of exercising animals indicates that both overproduction of sugar from protein and diminished utilization play a part in the creation of the diabetic state. The latter mechanism has certainly gained much support in recent years.

The respiratory quotient of the depancreatized animal not receiving insulin assumes the low level 0.69–0.73 and is not usually raised when sugar is given. Under certain conditions, in animals which have received a high protein diet and insulin, for a time the quotient may be higher (Soskin). The

characteristic low quotient may indicate combustion of fat but if any process, such as conversion of fat to sugar which gives a very low quotient, should be taking place the ratio might indicate the resultant between sugar combustion, fat combustion, and the conversion. In brief, since more than one interpretation of a quotient is possible great caution must be observed in drawing conclusions with regard to its significance. When the liver is removed from a depancreatized dog the quotient rises, indicating either that relatively more sugar is being burned or that conversions giving a low quotient are taking place to a smaller extent.

The glycogen content of the muscles of a depancreatized animal may be reduced below the normal level, but appreciable amounts remain. There is no diminution of heart muscle glycogen and indeed convincing evidence of an increase has been obtained. Liver glycogen falls to very low levels. A slight increase in both muscle and liver glycogen can be produced by giving large amounts of sugar without exogenous insulin. This might be due to small amounts of residual insulin which would now appear to persist, under certain conditions, for rather long periods.

While the glucose utilization of the normal heart is low, that of the diabetic animal is even less when the blood sugar concentrations are the same. Lactate, however, is used almost as well by the diabetic as by the normal heart (Lovatt Evans). These findings eliminate the necessity of supposing that the diabetic heart depends for its energy entirely on protein and fat. The rate of usage of glucose but not of lactate by the diabetic heart is increased when insulin is supplied.

The excretion of phosphorus is increased in the depancreatized animal. The administration of sugar or adrenaline does not cause the prompt fall in the inorganic phosphate of the blood which is observed in normal animals. These substances, therefore, produce their effects by raising blood sugar which in turn calls forth insulin. Adrenaline and insulin thus affect blood inorganic phosphate similarly, but insulin is effective in the absence of the adrenal glands, i.e. the action is in both cases due to the pancreatic hormone.

The diabetic animal is very susceptible to infections but it is not established that this is due to the raised sugar content of the tissues. Some diabetic patients have a decreased ability to form antibodies. This defect appears to be more closely related to hypoproteinemia than to hyperglycemia.

Pathological conditions are observed with considerable frequency in the eyes of diabetic animals. The liver rapidly undergoes extensive fatty degeneration and there may be an accumulation of large amounts of neutral fat.

While there is considerable variation in the length of life of the depancreatized dog or cat, most individuals fed on a mixed diet do not live for more than two or three weeks without insulin. Under certain conditions dogs may survive for seven weeks. It is now established, however, that when the anterior pituitary is also removed the animal (dog) may live for nine months, at least, and exhibit only a mild form of diabetes. Severe diabetes produced under these conditions by administration of the diabetogenic substances of the anterior pituitary is alleviated by insulin.

THE ANTIDIABETIC HORMONE—INSULIN

The name *insuline* was suggested by de Meyer in 1909 for the hypothetical internal secretion of the pancreas the search for which had been stimulated by von Mering's and Minkowski's findings (1889). While other workers, among whom Hédon, Zuelzer, and Scott may be mentioned, obtained very suggestive results, which in some cases were probably due to the presence of insulin, Banting and Best working in Macleod's laboratory (1922) were the first to obtain a preparation containing the antidiabetic hormone in a form which consistently alleviated all signs of diabetes in completely depancreatized dogs (fig. 49.1).

SOURCE OF INSULIN

While it is reasonable to suppose that small amounts of insulin are present in tissues other than the pancreas, methods are not yet available for their estimation. Blood provides an exception to this generalization but the active substance is detected by the administration of the whole blood and not as yet by extraction of the insulin from the tissue. In the mammalian organism the pancreas appears to be the only organ to manufacture insulin or to store it in more than minute amounts.

The islet cells of the pancreas are of four types— α , β , γ and δ (fig. 49.2). The α and β types contain granules. The γ are non-granular. The δ cells have been seen only in human pancreas (Bloom) and are not well defined. In dog pancreas the number of cells per islet varies greatly, as do the relative numbers of the various types of cells. One study gives the average number per islet as 30 and the average ratio of α to β cells as 20/75. The islet

volume may be about one one-hundredth of the pancreas. The β cells occupy the periphery of the islets and are smaller than the others. It is these cells which are considered to be producers of the antidiabetic hormone, indeed, the granules of these cells may consist, in part, of this substance. Epithelial cells of the small ducts are considered to be the "mother cells" of the islet and acinar cells. New islet cells may, therefore, be produced from them.

The main points of evidence which indicate that the hormone is produced in the islet cells are as follows:

1. Histologically, the islets are glandular structures, the obvious outlet for the secretion of which is through the blood stream.
2. There are relatively large amounts of the hormone in the principal islets of teleostean fishes, in which few enzyme-producing cells are found.
3. The active substance is found in degenerated pancreas in which the loss of acinous tissue has proceeded more rapidly than that of the islet cells. Ligation of the pancreatic ducts eventually produces a decrease in the insulin content of the pancreas, but moderate amounts of insulin may still be extracted when very few enzyme-producing cells remain.
4. When most of the pancreas, approximately nine-tenths, is removed from a dog, characteristic lesions (glycogen infiltration) are found in the β cells of the remnant. These changes can be accelerated by a high carbohydrate diet, and prevented or eliminated by administration of insulin or by fasting.
5. The clinical condition known as hyperinsulinism occurs when the pancreas liberates abnormally large amounts of antidiabetic hormone. In many of the cases there are definite tumors of the islet cells. After operative removal of these masses of islet cells the blood sugar is maintained at higher levels.
6. Metastases in other tissues arising from carcinoma of the islet cells have been shown to contain insulin.
7. The injection of anterior pituitary extracts leads to destructive changes in the islet cells, chiefly in the β cells, while there is little or no effect on the α cells.
8. Alloxan and other chemical substances destroy the cells of the islands of Langerhans, leaving the other cells essentially intact. The pancreases from a number of dogs treated with diabetogenic materials have been assayed for their insulin content, and the values obtained were roughly proportional to the concentration of granules in the β cells as determined by histological studies of these tissues. Good agreement between the amount of extractable insulin and the size and number of granules

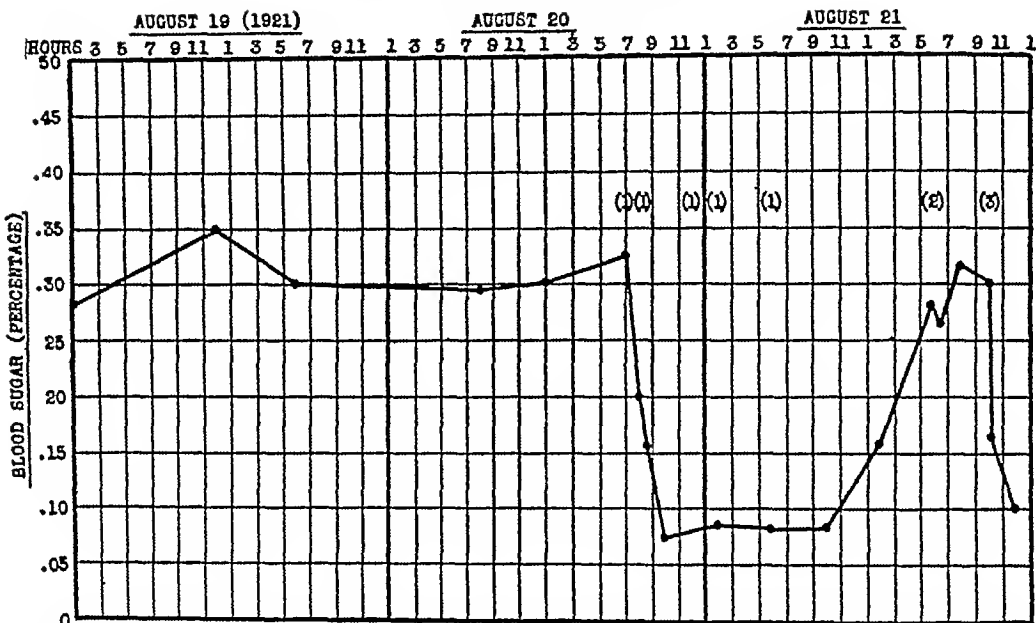


FIG 49-1 Effect of insulin on the blood sugar curve of a depancreatized dog (redrawn from Banting and Best) (1) Injection of extract of degenerated pancreas, (2) extract after incubation with pancreatic juice, (3) extract incubated without pancreatic juice. Blood sugars by Myers-Bailey modification of Lewis-Benedict method.

in the β cells of human pancreas has also been obtained

CHEMISTRY OF INSULIN

Insulin was first isolated in crystalline form by J. J. Abel in 1926. In 1934, D. A. Scott showed that it could be readily crystallized as the zinc salt and that Abel's crystals contained zinc. Nickel, calcium and cobalt also aid in effecting crystallization of insulin preparations. There is about 0.5 per cent of metal in the crystals. The "salts" of the insulin protein may appear in various forms—the zinc compound usually as twin plate-like rhombohedra. Chemical analysis by Brand and by the Cambridge workers under the leadership of A. C. Chibnall indicates that the insulin molecule is built up entirely or almost entirely of amino acids. Insulin is richer in the amino acids leucine, glutamic acid, and cystine than most other proteins, which are common in many proteins, are absent from the insulin molecule.

The maximum molecular weight of insulin is 48,000, pH 7.0–7.5, protein concentration 0.4–0.9 per cent. However, when more dilute solutions of the hormone are used below pH 4 or above 7.5 the insulin molecule dissociates into subunits having a molecular weight of about 12,000 (Gutfreund).

This is probably the physiological form of insulin although the physical chemists are actively debating the possibility of a smaller unit.

When insulin is oxidized with performic acid the molecule is split into its separate polypeptide chains. Two fractions can be isolated, A, an acidic fraction containing no basic amino acids and B, a basic fraction. These two physiologically inactive components are normally linked together by the —S—S— bridges of cystine and perhaps by other as yet unknown bonds. This work of Sanger's led to the conclusion that the 12,000 molecular weight insulin is composed of two identical A chains and two identical B chains. Sanger made a great contribution by introducing the use of partial hydrolysis of dinitrophenyl derivatives of the fractions to determine the sequence of the amino acid components. He had available the invaluable tool of paper chromatography of Martin and Synge and the newer knowledge of the chemical linkages attached by specific enzyme systems, such as pepsin, trypsin and chymotrypsin. The exact sequence of all the amino acids in the two chains of insulin is now known but we are still far removed from the knowledge of how the chains are formed and joined together in the beta cells of the islets.

When a dilute acidic solution of insulin containing a small amount of salt is heated, a flocculant

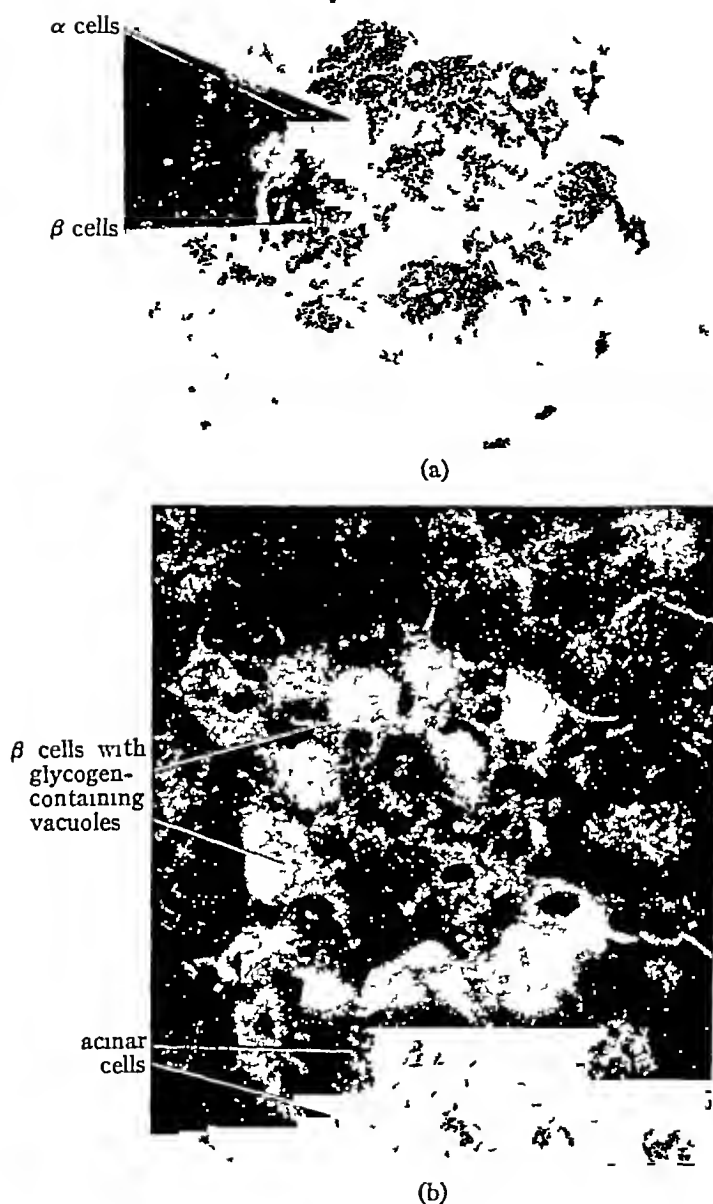


FIG 49 2

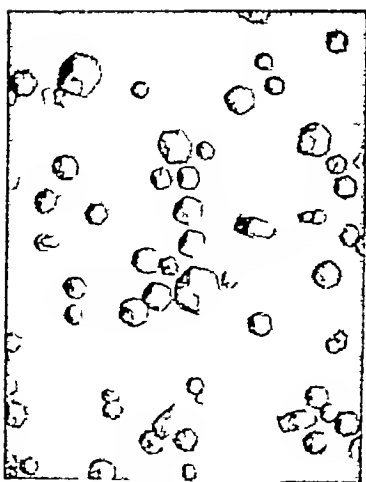
(a) Normal dog islet cells

(b) Islet from dog made diabetic with extract of anterior pituitary gland

precipitate forms. In experiments in which slightly different conditions were used, Waugh has shown that the insulin can be modified to yield fibrils. The rate of fibril formation increases with increasing hydrogen ion, salt and protein concentration, and with temperature. In fibril formation two reactions are involved. First the formation of active centers, and second the elongation of these into fibrils. These fibrils have little or no antidiabetic activity but can be converted into active insulin,

as can the so-called heat precipitate of insulin, by changing the reaction to the alkaline side. Seeding an insulin solution with fibrils may bring about a complete conversion of the active insulin into inactive fibrils.

Slightly acidified insulin has been kept for long periods, but in dilute alkali insulin is relatively unstable. Various attempts have been made to ascertain if there is in the molecule a specific grouping of certain of the amino acids which is really



Zinc Insulin Crystals

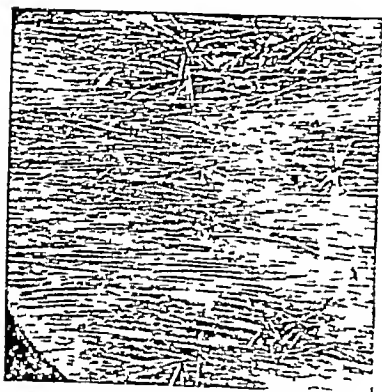
N amylamine Insulin Crystals
(Kindness of Dr D. A. Scott)

FIG 493

responsible for its hormonal activity. It may be concluded that the physiological activity of insulin may be slightly and sometimes reversibly decreased by certain minor chemical changes in the molecule, whereas appreciable chemical alteration gives a considerable diminution or complete absence of activity. Other changes have been suggested, such as the action of sulfhydryl groupings, as in glutathione, which may reduce the cystine disulfide linkage, a change which is known to be accompanied by inactivation.

STANDARDIZATION OF INSULIN

Zinc insulin crystals from all sources so far examined (man, cattle, hog, sheep, bison, fish) have the same potency. The international standard, a preparation of zinc insulin crystals is defined as

containing 24.5 units per milligram. There are two well established methods of assaying the potency of an insulin preparation. The lowering of blood sugar in fasting rabbits and the production of convulsions in fasting mice are both satisfactory effects of insulin for the comparison of unknown and standard products. This new, third, international yardstick of insulin is free of glucagon.

LIBERATION OF INSULIN

The arrangement of the capillary loops about the islet cells and the reported scarcity of lymph channels provide morphological evidence in favor of the capillary blood stream as the pathway by which insulin reaches the systemic circulation. It is important to remember that insulin passes first to the liver, but there is no proof that this tissue takes advantage of this favorable arrangement.

While there are many pieces of experimental evidence which support the conclusion that the level of blood sugar is an important factor in the regulation of insulin liberation, the possibility that a decrease in the rate of discharge of sugar from the liver may also be produced when the blood sugar is raised, is frequently overlooked. This latter effect is apparently produced by a direct action on liver cells. The injection of small amounts of glucose into the artery supplying a pancreas grafted into the neck of a depancreatized dog or into the pancreatic artery in a decerebrate cat causes a prompt lowering of blood sugar. In this latter case the effect was not obtained when the splenic or portal vein was used. The production of hyperglycemia, glycogen infiltration of the beta cells of the islets and permanent diabetes by the intraperitoneal administration of glucose (Lukens and Dohan) strongly suggests an overstimulation of pancreatic islets by the raised blood sugar level. Anderson and Long have demonstrated an output of insulin from perfused rat pancreas when dextrose was added to the perfusion fluid and Lukens and his group have obtained similar results in the dog. The results of experiments with denervated pancreatic grafts indicate that the nervous control is not essential. The nerve impulses which affect the islet cells are apparently conducted in part, at least, by the vagus. Vagus fibres have been traced to the islet cells and non-medullated branches are said to pierce them. The results of stimulating the vagus *appeared* to be clear cut and the pathway was traced by one group of investigators to the hypo-

thalamic region, but other workers have as yet been unable to confirm these findings. There may, of course, be a constant liberation of a small amount of insulin.

Information on the factors controlling the secretion of insulin has accumulated very slowly. The finding of Houssay, Foglia and their collaborators that two or three pancreases from normal dogs, when introduced in series into the carotid-jugular circulation of a depancreatized dog, produce no more effect than one pancreas, is strongly indicative of a chemical control of insulin liberation. These workers have studied the rate of liberation of insulin from the pancreas of hypophysectomized dogs, and of dogs made transiently or permanently diabetic by anterior pituitary extracts. The findings in general conform to those obtained by studying the insulin content of pancreas in relation to the state of carbohydrate metabolism of the animal (p. 676). Neither of these methods yields information as valuable as that which will be obtained when the insulin content of blood, and the rate of pancreatic blood flow can be accurately determined in intact unanesthetized animals.

INSULIN IN BLOOD

As a result of the work of Gellhorn, Anderson, Bornstein, Groen and their collaborators, methods for estimating the amount of insulin in blood are now available. These depend on the blood sugar response of hypophysectomized, adrenalectomized, alloxan-diabetic rats or mice or on the increase in sugar uptake of segments of isolated rat's diaphragm. While few laboratories have the necessary procedures readily available, some interesting results have been obtained. The value for normal human blood is approximately 4×10^{-6} mg per cc, i.e. $\frac{1}{250}$ microgram per cc. Some insulin can be detected in the blood of most maturity-onset diabetics but none in the blood of growth-onset, i.e. childhood, diabetics (Lawrence and Bornstein). These findings are in accord with Wrenshall's reports on extractable insulin from the pancreas in these two groups of diabetics. In the few cases of hyperinsulinism studied, a raised level of blood insulin has been found.

GENERAL EFFECTS OF INSULIN ON THE DIABETIC ORGANISM

It is well established that insulin restores to the depancreatized animal its ability to utilize sugars and fats in a normal manner. The excessive break-

down of protein is prevented. The ketosis rapidly disappears. Glycogen is deposited in large amounts in the liver. Muscle glycogen may be increased. The respiratory quotient rises when sugar is made available, or in fact when insulin alone is administered. Animals recover their ability to deal with infective agents. In brief, a well-treated depancreatized animal is difficult to distinguish from a normal one. There has been the difficulty, of course, that in the animal without a pancreas relatively large amounts of insulin are made available (by subcutaneous or intravenous injection) while in the intact animal small or large amounts are presumably liberated from the pancreas as the need arises. Adult depancreatized dogs on an adequate diet including the enzymes of the external secretion of pancreas have been maintained in good condition for more than six years (Macleod, Hédon, Bliss, Fisher, Wrenshall).

THE MECHANISM OF ACTION OF INSULIN

In both the diabetic and the normal animal the level of blood sugar is lowered by the administration of insulin. The sugar is not used in the blood itself but its rate of passage out into the tissues is accelerated.

GLYCOGEN STORAGE. Small amounts of glycogen may be deposited in the muscles and liver of the depancreatized dog when no insulin is given but the administration of insulin produces a dramatic increase in the rate of deposition in both these tissues. Stetten's work using deuterium as a tracer, indicates that the diabetic organism makes its liver glycogen from three carbon compounds (lactic acid) while the normal or insulin treated animal uses largely the six carbon compound (dextrose). In the normal animal one of the most clear cut effects of insulin is the increase in the rate of glycogen deposition in the muscle. It is obvious that insulin also promotes glycogen storage in the liver of the normal animal and this situation should not be confused with the results of adding more insulin to an already normal supply. In the normal adult animal there may be an actual loss of glycogen from the liver when extra insulin is administered. This may be due to the accelerated glycogen deposition in muscle, a process which has a high priority in the disposition of sugar, to the presence of glucagon in the insulin used, or to the increased rate at which fat is formed and sugar oxidized. These latter processes would decrease the amount of sugar available for storage as glycogen in the liver. The

liver slice is not a good preparation for the study of glycogen formation

The action of insulin can be well demonstrated in eviscerated and in isolated perfused preparations such as the hind limbs of the cat or dog. Dextrose must be supplied. Very great use has been made of the finding by Gemmell in 1940 that the rat diaphragm *in vitro* takes up more sugar and forms more glycogen when insulin is added. As little as 0.1 unit of insulin per kilo will double the diaphragmatic glycogen in the fasting adrenalectomized rat. Nelson (1944) demonstrated the inhibiting effect of a prior injection of anterior pituitary extract. The isolated rat diaphragm does not respond to insulin with an increased use of oxygen but the skeletal muscle from the same species does.

There is an increase in the glycogen of heart after removal of insulin from the body (Cruckshank). The renal tubules exhibit a deposition in the diabetic state which may be related to the increase in sugar which they are attempting to reabsorb. The beta cells in the islands of the pancreas show glycogen infiltration (Duff and Toreson) in the diabetic organism. This used to be termed hydropic degeneration.

ON FAT METABOLISM In the depancreatized animal insulin decreases the lipemia and cholesterolemia and prevents the deposition of the large amounts of excess fat which accumulate in the liver in the untreated animal. The level of the ketone bodies in the blood is restored to normal.

The fact that insulin increases the formation of fat has been obvious ever since the first emaciated dog or diabetic patient demonstrated a fine pad of adipose tissue, made as a result of treatment with the hormone. This effect was not perceptible in short term experiments without the aid of tracers and the disappearance of sugar was usually attributed to oxidation or perhaps in part to decreased gluconeogenesis. Using the technique of isotopic tracers, Stetten has calculated that in the well nourished rat only about 3% of the glucose ingested each day is converted to glycogen, while 30% is used to make fatty acids. In the absence of insulin the diabetic animal exhibits a much lower level of lipogenesis, as well as of glycogenesis.

In 1948 Bloch and Kramer were the first to demonstrate an *in vitro* effect of insulin on fat formation. Liver slices incubated in the presence of pyruvate and labelled acetate showed incorporation of the label into fatty acids and this

process was enhanced by prior administration of insulin. This work has been strongly supported and extended by Gurn with Brady and Lukens and by Chaikoff and his group. Thus we learn that insulin has widespread effects on fat formation and mobilization. It promotes lipogenesis from glucose and acetate in the liver and in many extrahepatic tissues. It prevents the loss of depot fat and the accumulation of fatty deposits in the liver. It suppresses ketosis by inhibiting the formation of ketone bodies. It is probable that these different effects of insulin are closely interrelated and that they are all linked with its action on the oxidation of glucose, i.e. lipogenesis is dependent on glycolysis.

ON PROTEIN METABOLISM The early work on insulin in Toronto demonstrated that insulin reduced the elevated nitrogen excretion characteristic of the diabetic state. In 1926 Janney and Shapiro noted a fall in blood urea and N.F.N. of normal human subjects given insulin and other investigators found a fall in blood amino acids in animals and man. In a classical study Mirsky showed that insulin decreased the rate of protein catabolism in the nephrectomized dog. Chaikoff and Forker in extension of F.N. Allan's original findings established a linear relation between insulin dose and nitrogen retention in depancreatized dogs. The evidence obtained by Forker, by Lotspeich, and by Sinex, MacMullen and Hastings indicates that insulin encourages the incorporation of labelled amino acids into protein. Krahil has recently reported the incorporation of C^{14} labelled glycine into glutathione and into protein in liver slices. This process is greatly reduced in diabetes but is restored to normal by insulin. The evidence suggests that insulin stimulates peptide synthesis rather than inhibiting breakdown and that this anabolic effect on protein metabolism is interrelated, perhaps dependent on, the effect of the hormone on sugar metabolism.

The literature records that Mirsky, Young, Gaebler, Lotspeich, Frame and Russell, and Lukens in distinct investigations have agreed that the presence of insulin is essential for the protein anabolic effect of the pituitary growth hormone, somatotropin. There is a great deal of indirect, but no direct, evidence that an increased secretion of insulin occurs in response to somatotropin in normal animals. Somatotropin has no anabolic effects in the complete absence of insulin but recent work has shown that insulin can stimulate

growth, i.e. nitrogen retention, bone growth, fat formation, etc., in the complete absence of somatotropin (Salter and Best)

ON OXIDATION OF SUGAR The interpretation of most of the early work on this subject was confused by the lack of a suitably labelled dextrose. In 1949 Villee and Hastings found that insulin increased the oxidation of glucose as judged by the rate of appearance of the carbon label in the CO_2 formed by the isolated rat diaphragm. Similar results have been secured by Sacks and Sinex who found that insulin increased the turnover of isotopic phosphorus in various phosphorus compounds, the deposition of C^{14} labelled glucose as glycogen, and the oxidation of glucose. Feller, Chaikoff et al., using C^{14} glucose have established that the rate of utilization was reduced to half the normal value in dogs by removal of insulin and restored completely by giving the hormone. Using C^{14} labelled glucose in eviscerated rabbits Wick, Drury, Bancroft, and MacKay were able to distinguish between the glucose oxidized and the total amount which disappeared. When insulin was given there was a delayed, but large, increase in the labelled carbon of the expired air and therefore in the amount of glucose oxidized. The rest of the labelled carbon was found in glycogen, fatty acids, proteins, and in non-glucose water-soluble compounds. Insulin did not affect the rate of acetate.

ON CELL PERMEABILITY The suggestion that a key point in the action of insulin may be on the permeability of cells to glucose has been revived and strongly supported by Levine and his colleagues. This view has been supported by Drury and Wick who interpret their evidence to indicate that insulin is not an oxidative catalyst and by Mirsky, who finds that entry into cells does not depend on the molecular configuration of the sugar as previous authors had suggested. It may be, therefore, that the principal site of insulin action is on the cell membrane, i.e. it merely accelerates the transfer of glucose into the cell and its subsequent fate depends on the state of the cell. The mechanism of the production of this apparent change in permeability—enzymatic or otherwise—will undoubtedly now receive great attention.

THE COMBINATION OF INSULIN WITH TISSUES Stadie and his colleagues have obtained evidence to support their working hypothesis that insulin and other hormones must form a rather firm chemical compound with the intact cells on which

they act. Isolated rat diaphragm dipped for a few seconds in a solution of insulin fixes insulin which reveals itself by increased utilization of sugar by the tissue. Diabetic tissue has a decreased ability to bind insulin. Growth hormone and cortisone inhibit this combination of insulin with muscle cells. Insulin injected into the living animal and much much higher concentrations in the fluid bathing the diaphragm *in vitro* lead to the binding of about the same effective amounts of insulin. Mammary and adipose tissue also bind insulin and this is perhaps a general phenomenon.

INSULIN AS A GROWTH HORMONE The various anabolic effects of insulin show that it possesses some of the characteristics of a growth hormone. True growth, including lengthening of the epiphyseal discs, can be elicited in completely hypophysectomized animals by giving gradually increasing amounts of slow acting insulin (Salter and Best). While it had been known that young hypophysectomized animals may grow for a time after removal of the pituitary (Selye and Collip), and while it had been suggested that growth hormone liberates insulin (Young), proof that hypophysectomized rats, which had in some cases not grown for months, could be made to resume growth by insulin had not been previously secured.

ON PHOSPHATE AND POTASSIUM Insulin produces a dramatic fall in the inorganic phosphate of blood plasma (Harrop). Ingestion of glucose or injection of adrenaline has the same effect but not in the absence of the pancreas. Levine, Loube and Weisberg (1949) have shown that the intravenous administration of fructose to untreated depancreatized dogs consistently produces a fall in blood inorganic phosphate. The change in phosphate is apparently a consequence of the rapid entry of hexoses into cells. Insulin is necessary for the entry of glucose but not of fructose. After insulin there may be a rise in the hexose monophosphate of muscle. This is an indirect effect apparently due to adrenaline liberation and a stimulation by it of phosphorylase action on glycogen and the formation of the first breakdown product, glucose-1-phosphate. Insulin does not significantly change the amount of adenosine triphosphate, phosphocreatine, or inorganic phosphate in muscle. There may be a rise in the A.T.P. fraction of liver. The rate of turnover or regeneration of these phosphate compounds is increased by insulin and this is probably an index of the increased rate of phosphorylation of sugar which is, as we have

seen, an indispensable preliminary step in its metabolism. The free energy for this change is made available by the conversion of adenosine triphosphate to adenosine diphosphate plus a high energy phosphate grouping. This phosphorylation of glucose then permits glucose to make available the free energy which it possesses. New molecules of A.T.P. are provided as a result of subsequent steps in the anaerobic breakdown of glucose.

There is evidence, as we have seen, that insulin may affect at least three of the enzyme systems involved in the metabolism of glucose. No further discussion will be attempted here on the action of insulin in stimulating glycogen synthesis (phosphorylase) or its possible effects in the citric (tricarboxylic) acid cycle except to stress that insulin definitely affects the utilization of acetate and pyruvate, i.e. steps distal to the hexokinase reaction.

There is a great deal of evidence, obtained largely in C. F. Cori's laboratory, that insulin stimulates the activity of hexokinase—the essentially irreversible first stage in glucose utilization. Growth hormone and cortisone inhibit this enzyme. Maximal activity as judged by the glucose uptake of isolated rat diaphragm is obtained when the pituitary and adrenal glands are absent and insulin is present in excess. In the diaphragms of animals in which only the insulin-producing cells have been destroyed a minimal rate of activity is observed. The hormonal control of hexokinase in tissues other than the diaphragm awaits further investigation.

Bornstein and Park in Cori's laboratory have found that serum from alloxan-diabetic rats inhibits the uptake of glucose *in vitro* by the diaphragm of normal fasted rats. Removal of either the adrenals or pituitary from the diabetic rat eliminates this inhibitory effect of its serum, and the injection of both growth hormone and cortisone restores the effect. These findings, therefore, suggest that the insulin-reversible inhibitor of glucose uptake is formed as the product of joint pituitary and adrenal activity. Studies will undoubtedly soon be made on the sera of diabetic patients.

Various changes in the concentration of the metallic constituents of the blood have been reported after insulin injection. These, and particularly that in the concentration of potassium, suggest a fundamental relationship between electrolyte and carbohydrate metabolism. The decrease

in blood sugar produced by insulin is accompanied by a simultaneous fall in potassium.

It is now well established that the administration of insulin lowers the insulin content of the pancreas in fasting or fed animals. Furthermore it protects the islet cells against the degenerative changes which occur after the removal of a large part of the pancreas. Similarly the degenerative changes in the islets and the loss of insulin content produced by the administration of the diabetogenic substance of the anterior pituitary gland are prevented if insulin is supplied. If, after the diabetic state has been produced in partially depancreatized cats by administration of diabetogenic preparations, insulin is given in adequate amounts, the diabetes may disappear. Thus insulin under these experimental conditions can both prevent and cure diabetes (Haist, Campbell and Best, Lukens and Dohan).

INTERFERENCE WITH THE ACTION OF INSULIN

Other hormones. There are five internal secretions the action of which may be considered to be antagonistic to that of insulin. There is no evidence of any chemical interaction of these hormones with the antidiabetic substance.

Enzymes. Insulin is destroyed by pepsin-HCl and by the activated proteolytic pancreatic enzyme, chymotrypsin. Crystalline trypsin does not destroy insulin. I. A. Mirsky and his colleagues have shown that the liver and to a lesser extent other tissues contain a factor which is capable of inactivating insulin and a second one which inhibits this action. These factors are referred to as insulinase and insulinase-inhibitor. Their physiological significance is not established but their presence suggests the possibility that an increased rate of destruction of insulin as well as a decreased formation must be considered. Pancreas may be incubated aseptically at neutral or acid reaction without loss of the antidiabetic substance.

Reaction of tissues. Since insulin acts on the cells of the liver and muscles, factors which influence these tissues, acidosis for example, may modify insulin action. Furthermore, since the liver is so largely responsible for the regulation of blood sugar, influences affecting this organ may cause a change in sugar content quite apart from the action of insulin. A change in the acid-base equilibrium of the body toward the acid side renders injected insulin less effective, a change toward the alkaline makes insulin more effective.

Products of infection The toxic products elaborated by many microorganisms may interfere with the action of insulin. There is experimental evidence (1) that the insulin content of pancreas is decreased in certain severe infections but this does not necessarily indicate a decreased rate of liberation of the hormone, (2) that the suprarenal and thyroid glands are stimulated to release more of their internal secretions,² (3) that the synthesis of glycogen from lactic acid in the liver is inhibited and (4) it now appears that normally liver glycogen is changed to glucose by glucose-6 phosphatase. Amylase, which splits glycogen to dextrins, maltose and finally to glucose, does not have access to the liver cells. When certain toxins are administered, however, amylase activity can be demonstrated and this abnormal route of glycogen breakdown provides another mechanism by which the products of infection make the organism resistant to the action of insulin. Insulin has no effect on the activity of amylase. Certain toxins may act on one or more of these mechanisms but investigation of this field is still in the preliminary stages.

Anesthetics All anesthetics interfere somewhat with the action of insulin. More or less asphyxia is produced by all general anesthetics. In asphyxia (1) adrenaline is liberated and (2) acid products tend to accumulate. Chloralose and amytal cause the least disturbance of carbohydrate metabolism.

SUBSTITUTES FOR INSULIN

Insulin therapy has two obvious disadvantages (1) the transient effect and (2) the necessity for parenteral injection of the active material. The most important of the suggested substitutes is synthalin (decamethylenediguanidine). This substance does not increase deposition of glycogen and the effect on hepatic gluconeogenesis is accomplished in a highly unphysiological manner. Extensive deposits of fat produced in diabetic dogs by fat feeding as well as by many chemical compounds other than synthalin interfere with sugar formation. It is preferable that the human diabetic organism should excrete some of the sugar made by a relatively healthy liver rather than to be made "sugar free" by damaging liver tissue so that less glucose is formed.

² The first effects of the products of infection may be to excite liver tissue to increased gluconeogenesis and discharge of glucose. Later the liver cells may be damaged so that less glucose is produced.

INSULIN REQUIREMENT AND ADMINISTRATION

Interesting studies have been made of the insulin requirements of depancreatized dogs under different conditions. Thus the blood sugar has been kept at a normal level by simultaneous and continuous intravenous injection of insulin and dextrose solutions. The insulin required was between 0.06 and 0.4 units per kilogram per hour, while the corresponding requirement of dextrose was 0.2 and 0.6 gm per kilogram per hour. The higher values for insulin and dextrose were those required by unanesthetized dogs, the others by anesthetized dogs. In another study the amount of insulin necessary to keep the blood sugar at a normal value in depancreatized dogs under basal conditions was between 0.005 and 0.035 units per kilogram per hour, with an average value of 0.017 units per kilogram per hour. The duration of action of insulin is not proportional to the size of the dose injected but is a simple function of the logarithm of the dose, i.e., insulin is inactivated in the body at a rate proportional to the amount in the body at the time. Thus if 1 unit lasts four hours, 10 units would last eight hours. It is established that the completely depancreatized human subject may require less insulin than many "spontaneous" diabetics. Similarly, the alloxan-diabetic dog may require more insulin than the depancreatized. There are three obvious possibilities to be advanced in explanation (1) The absorption of food is reduced by removal of the pancreas. Comparisons should be made on the same levels of nutrition. (2) The alpha cells of the pancreas elaborate an antagonist of insulin, glucagon. (3) The alloxan or the meta-pituitary diabetic may and frequently does have some pancreatic insulin available. This may be liberated more physiologically than exogenous insulin and thus help to maintain the organism in a condition which creates a demand for more insulin.

Insulin may be administered effectively by the subcutaneous or intravenous route. Some absorption may be obtained by inunction or by application to the sublingual or other mucous surfaces. Rectal administration is ineffective. Studies continue to be made on the problem of the oral administration of insulin. Efforts have been made to combine it with various materials, dyes, phenolic substances, tannic acid etc. which might protect the protein molecule from destruction by the intestinal enzymes. The difficulties involved are obvious, and it is therefore not surprising that,

while some success has attended these efforts in the laboratory, no satisfactory application to the treatment of diabetic patients has yet been made.

MODIFIED INSULIN

One of the most obvious difficulties in the use of regular insulin in clinical diabetes is its transient and sometimes too violent action. This difficulty has been much more frequently encountered since the highly purified preparations of insulin have been made available. The cruder products were absorbed more slowly. While a great many attempts have been made to slow and prolong the action of insulin, the first important success has been obtained by Hagedorn, Jensen, Krarup and Wodstrup. These investigators have shown that a compound of insulin with any one of several protamines exerts a slower and more prolonged anti-diabetic effect than regular insulin. This has been shown by microscopic observation (Beecher and Krogh) to be due to the much slower absorption of the insulin combined with protamine than is the case with the regular preparations of this substance.

Completely depancreatized dogs may be maintained sugar free, while receiving a very liberal diet, on one dose of protamine insulin daily without the development of any hypoglycemic reactions (Kerr and Best).

Scott and Fisher demonstrated that protamine insulin is greatly improved by the addition of a small amount of zinc. The resulting product, protamine zinc insulin, exerts a more prolonged hypoglycemic action and forms a much more stable suspension than protamine insulin.

Various other forms of modified insulin have been prepared and tested clinically. Histone insulin and globin insulin are examples. Protamine zinc insulin has been prepared in crystalline form (Hagedorn) and free insulin is stable when added to this compound. Zinc insulin preparations, without protamine, exerting a prolonged action, have recently been made available by Hallas-Møller et al.

HYPOGLYCEMIA

Under certain exceptional circumstances hypoglycemia may be produced by excessive utilization of glucose (prolonged very violent muscular exercise) but interference with the formation of sugar in the liver is largely responsible for most types. The three main factors which diminish sugar pro-

duction may be classified as follows: (1) abnormality of liver cells, (2) the inhibiting action of insulin on gluconeogenesis in the liver, and (3) the decreased hepatic gluconeogenesis resulting from diminished output of the anterior pituitary, thyroid, or cortical and medullary adrenal secretions. Under the first heading a great variety of experimental and clinical conditions may be listed—for example—phosphorus or hydrazine poisoning, yellow fever, acute yellow atrophy, and the bacterial infections. When the normal liver is completely removed, profound hypoglycemia occurs promptly. Approximately 80 per cent of the normal liver must be removed before hypoglycemia is produced. Under the second heading we may consider *hyperinsulinism*. This term should be reserved for conditions in which it is established that there is liberation of excessive amounts of insulin from the pancreas. This has been the case in numerous instances in which the removal of a tumor of islet cells has corrected the hypoglycemia. Correction of the condition by removal of a large part of the pancreas does not prove that the cause was liberation of abnormal amounts of insulin since decreasing the amount of insulin may merely compensate for the first abnormality. The relative importance of the three glands of internal secretion listed under the third heading may vary in different species. Removal of the thyroid increases the sensitivity of an animal to insulin and the same is true of the adrenal medulla, but hypoglycemia is not produced. When the anterior pituitary is extirpated, however, there may be profound hypoglycemia and this finding suggests that diminished secretion of the diabetogenic substance may be an important factor in certain clinical cases in which the liver and pancreas appear perfectly normal. In some species removal of the whole adrenal gland causes hypoglycemia and this is, at least partially, corrected by the administration of cortical extract. Clinically certain cases of Simmonds's disease (diminished anterior pituitary secretion) and of Addison's disease (involvement of adrenal cortex) may exhibit hypoglycemia. (See table 63.)

SIGNS AND SYMPTOMS

The signs of hypoglycemia were first adequately described by Mann and his collaborators. The low blood sugars were produced in dogs by removal of the liver. The description of this condition enabled the Toronto investigators to recognize that the effects of large doses of insulin were the same as

TABLE 63
Spontaneous hypoglycemia or dysinsulinism

<i>Hyperinsulinism* (Hyperactivity or tumor of islands of Langerhans)</i>	<i>Interference with gluconeogenesis in liver</i>	<i>Hypofunction of an- terior pituitary, adrenals or thyroid</i>
<i>In experimental animals</i>	<i>In experimental animals</i>	<i>In experimental animals</i>
Hypertrophy after duct li- gation?	Hepatec- tomy	Removal of pituitary (anterior lobe)
Hypertrophy after anterior pituitary ex- tracts?	Interference with arte- rial blood flow	Removal of adrenals
No tumors care- fully studied	Poisoning Phosphorus Chloro- form Hydrazine Synthalin, etc.	Removal of thyroid Hypogly- cemia or (in- creased sus- ceptibility to insulin)
	Deposition of fat	
<i>Clinical observa- tions</i>	<i>Clinical observa- tions</i>	<i>Clinical observa- tions</i>
Cases of hyper- plasia and hy- pertrophy	Hepatitis Carcinomata Yellow fever	Hypogly- cemia in so- called pitu- itary ca- chexia or Simmond's disease
Cases of tumor	Acute yellow atrophy Poisoning Phosphorus Carbon tet- rachloride Benzol Chloro- form Synthalin, etc.	Hypogly- cemia in Ad- dison's dis- ease (some cases)
	Surgical in- terference with blood flow	Increased sus- ceptibility to insulin after thy- roidectomy

There may also be excessive utilization of sugar by muscles—dogs in tread mill, marathon runners, etc

* The term "hyperinsulinism" was introduced by Seale Harris in 1924 to describe cases exhibiting signs of hypoglycemia. The first report of a tumor of the islet cells which in this case had secondary growths in the liver was made by Wilder, Allan, Power and Robertson in 1927. Many cases of tumor have now been studied. A histological section from the first tumor successfully removed is shown in fig 49 4

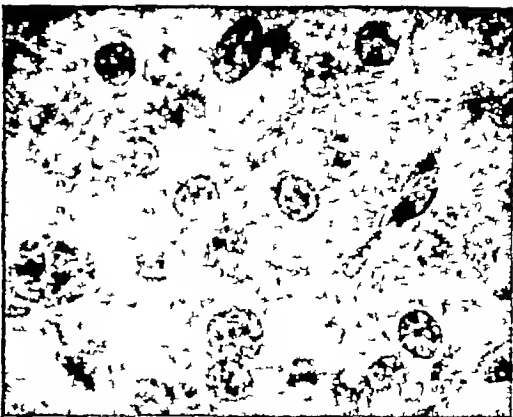


FIG 49 4 Photomicrograph of adenoma of pancreatic islets associated with clinical hypoglycemia. This tumor illustrated was the first to be surgically removed. Hematoxylin and eosin stain $\times 1000$. From Howland, G, Campbell, W R, Maltby, E J, and Robinson, W L, J A M A, 93 674-79, 1929

those due to hypoglycemia produced by other means. The signs and symptoms vary in the different species. The first signs in the rabbit are hyperexcitability and desire for food. The excitability becomes greater, and mild and then severe convulsions are exhibited. The head is retracted and the hind limbs extended in the intervals between convulsive seizures. Coma is frequent. The animals may exhibit rigor mortis immediately after death. The signs in dogs are quite similar. Mice, in some instances, may become comatose without exhibiting convulsions. Cold blooded animals do not show any signs until many hours or even days after insulin injections³. The signs and symptoms in man have been extensively studied in the laboratory and clinic. The initial symptoms may be hunger or a feeling of nervousness—a sense of impending danger. A little later there may be profuse perspiration, alternate pallor and flushing of the face, vertigo and diplopia. The blood sugar at this stage is 0.06-0.04 per cent but the level varies greatly in different individuals. Most hypoglycemic reactions proceed no further than this. In very severe cases there may be delirium, convulsions, and death. The true blood sugar may decrease until only non-sugar-reducing power remains.

Sakel's insulin-shock treatment for schizophrenia has focussed attention on the metabolism of brain and the effects of prolonged hypoglycemia. Brain tissue utilizes carbohydrate almost exclusively. Hypoglycemia interferes with the supply and pro-

³ Similar prolonged delay may be observed in mammals after injection of huge doses of insulin.

duces much the same condition as O_2 deficiency. The electrical activity of the cerebral cortex is depressed in hypoglycemia and restored to normal by the administration of glucose. The reduction in the oxidative metabolism of brain is undoubtedly responsible for this and other changes.

Cerebral damage, which may be permanent, has been observed in both animals and man as a result of prolonged hypoglycemia.

ALLEVIATION OF HYPOGLYCEMIA

The intravenous administration of glucose is the most effective method of alleviating hypoglycemia. The prompt recovery of almost moribund animals provides one of physiology's most fascinating demonstrations. Mannose is almost as useful as glucose, and fructose also occupies a preferential position. Galactose and maltose have a slight but transient effect. Sucrose,⁴ lactose and pentoses are not effective. Glycogen and glycerol have been shown to exert some beneficial action. The effect of these substances in hypoglycemia is probably largely dependent on the rapidity with which they are transformed into glucose in the liver. Fructose may be converted slowly to glucose in muscle but there is a possibility that it may be burned directly (Griffiths and Waters). It is generally assumed that the only sugar directly oxidized in the muscle, where a large part of the total oxygen use takes place, is d glucose. The usefulness of the other carbohydrates would therefore depend on the ease of their conversion into this sugar. Adrenaline and pituitrin may be used to alleviate hypoglycemia but glucose is much more efficacious and safe. Liberation of adrenaline is, however, one of the physiological mechanisms by which hypoglycemia is corrected. Intense anger, such as that which might well be experienced by a diabetic whose hypoglycemia was mistaken for alcoholic intoxication (Duncan) may correct hypoglycemia through liberation of adrenaline.

THE INSULIN CONTENT OF THE PANCREAS UNDER DIFFERENT CONDITIONS

The insulin content of the pancreas has been determined in various animal species. The insulin is extracted from minced pancreases with an acid aqueous alcohol solution. Certain contaminating material is removed and the active material is

⁴ Carbohydrates which form glucose are, of course, effective. The above statement refers to the results of intravenous injections.

precipitated. This is then redissolved and estimated by the mouse method of assay. In the dog the insulin content of the free splenic end of the pancreas is greatest, that of the attached duodenal portion has an intermediate value, while that of the free duodenal end is lowest, the values being about 4, 3, and 2 units per gram, respectively. In partially depancreatized dogs, provided sufficient pancreas is left to prevent the onset of diabetes, the insulin content does not differ from that of the corresponding part in a normal dog, nor are any degenerative changes in the β cells noted. If diabetes supervenes, hydropic degeneration of these cells is observed, and the insulin content of the remnant of pancreas falls to extremely low values. The daily injection into dogs of diabetogenic extracts from the anterior lobe of the pituitary gland produces a prompt and profound decrease in the insulin content of the pancreas (in seven days to 0.2 units per gram). If the injections are stopped at this stage, the insulin content is restored to normal within four days. If the administration is continued the insulin is reduced to negligible amounts. No recovery will occur when this point is reached. Simultaneous administration of insulin prevents or greatly modifies the fall in the insulin stores. This fact strongly suggests that the β cells are permanently damaged by the extract through overwork and that the simultaneous administration of insulin relieves the cells of some of this excessive demand for the hormone.

Starvation (seven days) or a diet rich in fat produces a decrease in the insulin content of the rat's pancreas to about half the normal value, which is about 2½ units per rat. These animals have their insulin stores speedily restored to normal (in six days) when they are returned to a balanced diet, carbohydrate alone effects a partial restoration. Daily injection of insulin into rats causes an even more marked decrease in the insulin content of the pancreas than does starvation (Haist and Best). Massive doses of insulin over prolonged periods may produce atrophic changes in the pancreatic islets of partially depancreatized dogs which survive the treatment (Mirsky).

The injection of anterior pituitary extracts in certain strains of rats increases the islet volume and the insulin content of pancreas (Young, Richardson and Marks). In other strains there is no increase in insulin content.

The factors which affect the volume (growth) of the β cells of the pancreas are given in table 64.

Subcutaneous estrogen transplants or the administration of stilbestrol produce an increase in the insulin content of rat's pancreas. This effect is not observed in the absence of the pituitary (Griffiths, Marks and Young, Funk.) Stilbestrol may exert a diabetogenic effect in force-fed normal and partially depancreatized rats, and under these conditions Ingle has been able to demonstrate diabetogenic effects of all the estrogens. Under more physiological conditions Houssay, Foglia, Martinez and their colleagues have demonstrated that the estrogenic substances decrease the incidence of diabetes in partially depancreatized rats—presumably by stimulation of the β cells in the pancreatic remnant.

The effect of age on the insulin content of the pancreas has been studied in the cow. In fetal calves under 5 months the concentration was 34 units per gram, in calves 6 to 8 weeks old, 10 units per gram, in heifers 2 years old, 5 units per gram, in cows over 9 years, 2 units per gram. Pregnant cows 7 years old and older showed no change from the normal insulin content of about 2 units per gram (Scott and Fisher). In Wistar rats the total insulin content of the pancreas increases with age.

Pancreases obtained from non-diabetic persons at autopsy have an average insulin content of about 2 units per gram. This is probably somewhat lower than the true value. Pancreatic tissue from diabetic persons shows wide variations. The pancreas from diabetic children contains very little insulin. Pancreas from diabetic adults contains on the average some 40 per cent of the normal amount of insulin (Wrenshall). The insulin content of a tumor of islet tissue surgically removed from a patient suffering from hyperinsulinism may be as high as 214 units per gram.

It is, of course, apparent that these "insulin contents" indicate the balance between the rate of production of the hormone in the islets and the rate of liberation. There is good reason to believe that under certain conditions the rate of liberation is proportional to the content. Under other circumstances this is probably not true, but the rate of liberation of insulin can at present only be inferred from data which are susceptible of various interpretations. The conclusion has been drawn from some of these results, combined with histological studies, that the islet cells are "rested" after administration of insulin, by starvation and by a high fat diet and that less insulin is excreted by the pancreas than under normal conditions. Partial

TABLE 64
Factors depressing islet growth

- 1 Restriction of caloric intake
Restriction of carbohydrate intake
- 2 Administration of large amounts of insulin
- 3 Removal of the pituitary gland

Factors increasing islet growth

	INTACT	HYPOX
1 High carbohydrate intake		
2 Continuous injection of glucose		
3 Injections of anterior pituitary extract	+	+
4 Injections of growth hormone preparations	+	+
5 Injections of ACTH	+	+
6 Injections of cortisone	+	+
7 Thyroid administration	+	+
8 Estradiol benzoate	+	†
9 Diethyl stilbestrol	+	†
10 Progesterone	+	†
11 Testosterone	*	†

From R. E. Haist

* Not significant

† Not done

pancreatectomy, sufficiently extensive to result in diabetes, or administration of diabetogenic extracts causes (1) marked stimulation of the islets and (2) subsequent degenerative changes and loss of insulin.⁵

THE USE OF INSULIN IN NON-DIABETIC CONDITIONS

Favorable results have been claimed for the use of insulin in a very great variety of non-diabetic conditions. It has been used in pernicious anemia, in acute infectious diseases, in eclampsia, in pernicious vomiting of pregnancy, and in hepatitis—to mention only a few. While it is conceivable that insulin might be of slight benefit in some of these conditions, it would appear that equally satisfactory results can be secured by the administration of glucose alone. It is a clinical fact that the administration of glucose produces favorable results in a variety of hepatic abnormalities. A high glycogen content appears to protect the liver cells from damage and inhibition of gluconeogenesis produced by both insulin and glucose may also play a role.

⁵ For references to work on insulin content of pancreas, see Haist, 1944 and Wrenshall, 1952.

Insulin can now be considered, however, an established adjuvant in the treatment of certain cases in which lack of appetite prevents the ingestion of adequate amounts of food. The physiological basis for this use of insulin in these non-diabetic individuals rests very largely upon the increase in hunger and appetite which may be caused by the administration of sufficient material to produce a definite but not too marked hypoglycemia. In 1924 Bulatao and Carlson reported that production of hypoglycemia in experimental animals by the subcutaneous injection of insulin was uniformly accompanied by hypertonus and hypermotility of the stomach. The gastric tonicity and motility increase as the hypoglycemia deepens until complete tetanus is reached, which persists until the dog exhibits hypoglycemic convulsions or until sugar is given. The effect of sugar is immediate, but if a large dose of insulin has been given the hyperactivity of the stomach returns as soon as the blood sugar falls again. The first record of the increase in gastric peristalsis in the human individual after insulin administration was that of Dickson and Wilson, 1924. An hour and ten minutes after the administration of insulin the tone, depth and rate of peristalsis and rate of emptying of the stomach were definitely increased. This condition persisted for two hours, when the blood sugar was found to be 70 mg per 100 cc. Glucose was then given, and while the acute symptoms were definitely relieved the hunger persisted. Subsequently Quigley, Johnson and Solomon studied the effect of insulin on the gastric movements of four human subjects. They found that doses of from 12 to 20 units of insulin definitely increased the gastric activity. The first definite augmentation was observed about an hour after the injection and persisted for at least five hours. There was prolonged duration of the hunger period, and this was considered to be the most characteristic effect of insulin. The increased peristalsis produced by insulin is not inhibited by such procedures as smoking, unpleasant emotions, body discomfort, or the presence of moderate amounts of non-carbohydrate food in the stomach. The increased movement was inhibited by atropine. The immediate relief of the excess peristalsis when appropriate amounts of glucose were given was confirmed. More recently Grossman and Stein have shown that the sensations of hunger induced by insulin, with the exception of the epigastric pangs of distress associated with individual gastric contractions, continue to occur after complete vagotomy in man.

It is known that insulin does not exert its effect on gastric motility and secretion after section of the vagus nerves. This means that either insulin acts centrally by stimulating the vagus, or that the continued elaboration of acetylcholine, which we now know is an essential part of the mechanism by which the vagus exerts its action, provides a foundation upon which the peripheral effects of insulin may be superimposed.

Insulin augments also, to some extent, peristaltic movement in the duodenum and in the colon, but the effect is not as marked as in the case of the stomach. It will be remembered that one of the early symptoms observed in experimental animals and also in human subjects after the administration of insulin is an increase in hunger. In animals an attempt to consume material of little nutritive value which under ordinary conditions they would not attempt to eat is often observed. It would appear, therefore, that the clinician is able to take advantage of this situation by providing nutritious food for his hungry patient. The increase in weight observed in both animals and patients when appropriate doses of insulin are given for prolonged periods is due to increased deposition of fat, and to a lesser extent to increased deposition of carbohydrate and protein. The increase in weight is not attributable to unphysiological retention of water.

INFLUENCE OF OTHER ENDOCRINE GLANDS ON CARBOHYDRATE METABOLISM

THE ADRENALS

The glucosuria produced by adrenaline was first noted by Blum. The intravenous route of administration gives the greatest rise of blood sugar but subcutaneous, intramuscular or intraperitoneal injections are effective. The immediate rise in the sugar of the blood is due to breakdown of liver glycogen to glucose. There may be a considerable decrease in the amount of liver glycogen. Adrenaline also mobilizes muscle glycogen (Cori) but here the immediate product is lactic acid and not glucose. A part of the lactic acid is carried by the blood to the liver where it is converted to glycogen, which in turn furnishes the blood with glucose. Muscle glycogen is therefore available indirectly to replenish blood glucose. When lactic acid from the muscles has been changed in appreciable amounts to glycogen in the liver the amount of substance in this organ may be increased over the normal level. Adrenaline therefore in moderate dosage, first causes a decrease and then an increase in liver glycogen. Very large doses over prolonged periods may lower both muscle and liver glycogen. The nervous control of adrenaline secretion will be discussed in chapter 59, and the manner in which thoracic autonomic impulses may affect blood sugar through its liberation will be appre-

ciated Adrenaline does not accelerate sugar formation from other substances in the liver Diabetes has not been produced by the continued administration of adrenaline

Insulin and adrenaline are not chemical antagonists but possess opposing physiological actions Adrenaline accelerates the breakdown of both liver and muscle glycogen but the lactic acid made from muscle glycogen may result in an actual increase in the liver Insulin promotes the formation of glycogen in both organs but the increase in muscle glycogen may be at the expense of sugar, which would have formed liver glycogen These are excellent examples of the manner in which the action of a hormone may be obscured by other effects When the blood sugar is lowered to about 0.06 per cent by insulin an increased rate of liberation of adrenaline may be detected (Cannon, Houssay) Adrenaline has also been reported to increase the rate of liberation of insulin from the pancreas

Adrenaline is one of the steps in the mechanisms by which liberation of adrenocorticotrophic hormone (ACTH) is released in animals (Vogt, Long and Fry) but it has not been possible to demonstrate this effect in man Thus a substance released in emergency causes a prompt hyperglycemia and may at the same time set in motion other changes (liberation of ACTH) which stimulate gluconeogenesis and thus provide for the continuation of the blood sugar rise On the other hand, a meal high in carbohydrate given to fasted rats causes a reduction in the cholesterol content of adrenal cortex (liberation of cortical hormone) which is coincident with a rise in liver glycogen (Abelin) This and related studies are perhaps the beginning of a more complete understanding of the physiological coordination of medullary and cortical activities There is some evidence that a mechanism even more rapid than one which could be mediated by adrenaline may play a role in ACTH liberation

THE ADRENAL CORTEX AND CARBOHYDRATE METABOLISM After double adrenalectomy in some species (cat and rat) the carbohydrate reserves of the liver and muscles are depleted and there may be definite hypoglycemia (Britton and Silvette) These workers showed that the hypoglycemic condition may be corrected when the cortical hormone and glucose are provided but not when the latter alone is given They reported that the administration of extracts of the cortex elevated the carbohydrate stores of adrenalectomized animals well beyond the normal limits Insulin and glucose do

not increase glycogen deposition in the adrenalectomized animal unless cortical extract is supplied (Britton) but it is possible that this effect might occur if the animals were supplied with a diet adequately low in potassium and high in sodium Long and his collaborators have greatly extended the earlier experiments of Britton and Silvette and find rather remarkable increases in liver glycogen in fasted or fed normal animals when extracts of the adrenal cortex are administered Muscle glycogen was not affected by the cortical material in these experiments but may be increased when adequate amounts of sugar are supplied

After Houssay's demonstration that removal of the pituitary gland attenuates the severe diabetes resulting from total pancreatectomy in the toad and dog, Long and Lukens showed that a very similar change could be produced by adrenalectomy in the depancreatized cat or dog It has been shown by Lukens and Dohan that the diabetes of adrenalectomized depancreatized animals and also that of hypophysectomized depancreatized animals can be increased in severity by the injection of cortical extracts

Long, Katzin and Fry have shown that the rat, partially depancreatized by the method of Shapiro and Pincus, is an excellent preparation for the demonstration of the role of the adrenal cortex in carbohydrate metabolism Adrenalectomy attenuates the diabetes which may be observed in these animals The grafting of cortical tissue may restore the glucosuria to the extent which has been observed before adrenalectomy

The relative diabetogenic potency of various adrenal steroids has been studied by Long and his collaborators, by Kendall and by Ingle and Thorn In 1941 Ingle demonstrated adrenal steroid diabetes in normal rats which were forced to ingest a diet rich in carbohydrate but of normal caloric value This adrenal diabetes was highly resistant to insulin Corticosterone, 17-hydroxycorticosterone and 17-hydroxy-11-dehydrocorticosterone have all been shown to produce hyperglycemia and glucosuria in the normal rat Ingle, Li and Evans have induced these signs of diabetes in normal force-fed rats by the administration of pure adrenocorticotrophic hormone The hypertrophy of the adrenal cortex was very marked in these animals The production of a decreased carbohydrate tolerance in human subjects by ACTH (Browne) and of a state of metabolism similar in many respects to clinical diabetes by administering more highly purified adrenocorticotrophic hormone (Conn, Louis and Johnston) are landmarks in the

clinical extension of laboratory findings. The diabetes in human subjects which persisted for as long as the injections were continued (10 days) was resistant to insulin. (See also Alloxan Diabetes, p. 685.)

It has been shown by Long and his collaborators and by Ingle and other workers that the mechanism through which these adrenal extracts produce glucosuria is in part by stimulating gluconeogenesis from protein, but as the increase in nitrogen excretion is insufficient to account for the extent of the glucosuria, interference with glucose oxidation has also been postulated.

Some but not all of the abnormalities of carbohydrate metabolism in adrenalectomized animals are apparently due to the disturbance in sodium and potassium metabolism. Thus the delayed absorption of sugar and fat and the failure to store glycogen from glucose can be favorably affected by appropriate salt treatment. On the other hand the sharp fall in carbohydrate levels in fasting adrenalectomized animals and the amelioration of diabetes in partially depancreatized rats by adrenalectomy are not corrected by feeding salt but are by the administration of suitable cortical material.

While it has been shown that the pituitary preparations may exert a diabetogenic action in the absence of the adrenal a part of the pituitary effect is undoubtedly exerted through the adrenal cortex. The recent work of de Bodo has demonstrated that the sensitivity of the hypophysectomized dog to insulin is considerably greater than that of the adrenalectomized dog.

Many of these findings which have been obtained on experimental animals have been confirmed by Thorn and his colleagues in studies on patients with Addison's disease.

THE PITUITARY AND CARBOHYDRATE METABOLISM

It has been appreciated for many years that abnormalities in carbohydrate metabolism may be associated with acromegaly of long standing or with the presence of various pituitary tumors. In 1908 Borchardt found that an extract of the posterior pituitary raised the blood sugar. In 1911 Cushing observed that "pituitary deficiency" may be accompanied by an increased carbohydrate tolerance. When insulin became available Burn demonstrated an antagonism between "pituitary" and insulin, and several workers (Olmsted, Geiling, Houssey) showed that animals were more sensitive to insulin after removal of the pituitary body.

It appeared for a time that both the oxytocic

and pressor principles of the posterior lobe produced hyperglycemia and thus interfered with the action of insulin. The work of Ellsworth and others indicates that the oxytocin is probably the more important factor from this viewpoint but it is doubtful, from the dosage necessary to produce the effect, whether this is of physiological significance. Griffiths has reported that posterior lobe extract, i.e., the vasoconstrictor principle, interferes with the absorption of insulin. The action of subcutaneously administered insulin is inhibited but when the intravenous route is used this effect is not observed.

Interest has been focussed on the anterior lobe of the pituitary by the brilliant researches of Houssey and his collaborators and of other investigators. The main points established in Houssey's laboratory are as follows: (a) Removal of the anterior lobe of the pituitary increases the sensitivity to insulin of the normal animal and diminishes the intensity of diabetes in the depancreatized animal. (b) Injections of preparations from the anterior pituitary into normal or hypophysectomized animals diminish their sensitivity to insulin and increase the severity of the diabetic state in hypophysectomized-depancreatized (Houssey) animals. (c) The administration of a suitable extract of the anterior pituitary can induce a diabetic condition. This point was independently established by the reports of Evans and his colleagues and of Baumann and Marne.

The effects of pancreatectomy and hypophysectomy are contrasted in the following summary, and the condition of the animal from which the pancreas and the pituitary have both been removed is briefly described.

<i>Pancreatectomy (D-1)</i>	<i>Hypophysectomy (D-1)</i>
Hyperglycemia	Low blood sugar, hypoglycemic convulsions during fasting
Polyuria	
Glycosuria	
Ketonuria	
Azoturia	
Insulin necessary for survival	Greatly increased sensitivity to insulin
Metabolic rate normal or slightly raised	Low metabolic rate
Decreased ability to utilize carbohydrate	Carbohydrate furnishes increased proportion of fuel
Decreased ability to form glycogen from fat and protein	Rapid disappearance of liver and muscle glycogen due to utilization of carbohydrate and decreased gluconeogenesis. ⁴

⁴For direct evidence of this effect see Crandall and Cherry.

Pancreatectomy and hypophysectomy

Animals survive without insulin Polyuria, glucosuria, ketonuria, azoturia slight or absent Administered carbohydrate partially or completely retained, i.e. carbohydrate utilization much better than in depancreatized dog Metabolic rate low Glycogen deposition

PERMANENT DIABETES The anterior pituitary gland contains a number of substances which affect carbohydrate metabolism in a variety of ways These will now be considered briefly Evans, Meyer, Simpson and Reichert in 1932 demonstrated the production of a prolonged diabetes in normal intact animals by anterior pituitary extract F G Young, 1938, was however, the first to produce a permanent diabetes comparable in intensity to that resulting from complete pancreatectomy, by injection of extracts of the anterior lobe of the pituitary He has found, in a very large number of dogs, that he is able consistently to produce a permanent state of diabetes by the daily injection, either intraperitoneally or subcutaneously, of a preparation of anterior lobe material The permanent state of diabetes may be produced after as few as eleven daily injections but more are usually required The diabetogenic activity, that is, the active material which will produce permanent diabetes in dogs, is associated with the globulin and pseudoglobulin fraction of the pituitary extract Young's work was confirmed by Campbell and Best and by Dohan and Lukens and many others Degenerative lesions of the islet cells of the pancreas were first noted in these permanently diabetic animals by Richardson and Young Signs of proliferative changes in the islet cells in the early stages of the injections were found by these workers and by Ham and Haist who also observed proliferative changes in the acinar and duct cells of the pancreas The diabetogenic extract produced proliferative changes in various other glandular tissues in the body Campbell, Haist and Best noted that the diabetic state produced by the pituitary extract was not intensified by complete removal of the pancreas and secondly, that the insulin content of pancreas was reduced to a negligible quantity This latter point has been discussed elsewhere

It would appear that the main effect of the substance or substances which produced the permanent diabetes is exerted upon the Islands of Langerhans These cells are apparently first stimulated and then destroyed by the repeated injections of the active material

The permanent diabetes produced by the above procedures differs from that caused by pancreatec-

tomy, in that the animals may live for long periods without the administration of insulin In some cases, however, insulin is required Starvation or a diet very rich in fat causes a marked diminution in the intensity of diabetes in the permanently diabetic animal

The production of permanent diabetes by the diabetogenic substance of the anterior pituitary can be prevented by the simultaneous administration of large doses of insulin (Haist, Campbell and Best) The islet cells are protected from profound degenerative changes, the insulin content of the pancreas remains at a moderately high level, and the state of permanent diabetes is not induced Lukens and Dohan have shown that permanent diabetes produced in partially depancreatized cats by administration of the diabetogenic substance of the anterior pituitary gland can be cured by the use of insulin, by a reduction in the caloric value of the diet or by an increase in its fat content Recovery from early diabetes has followed a reduction in the diet only when the diabetes was very mild but treatment with insulin produced recovery at this stage regardless of the severity of the disease If treatment were delayed until after the Islets of Langerhans had become atrophic, no recovery was possible

It is to be noted that in the prevention and cure of this experimental diabetes, the level of the blood sugar is probably one of the most important factors which determines the direction in which the islet lesions will progress This interpretation is supported by all the results which Young, Haist, Campbell and Best, and Lukens and Dohan have obtained and is in line with the earlier studies of F M Allen, Copp and Barclay, and others The rise of blood sugar in itself would not, however, constitute an adequate stimulus for the production of the extensive changes in the islet cells The diabetogenic factor must operate through other mechanisms as well The level of blood insulin may be one of these

Highly purified growth hormone has been shown to be diabetogenic, i.e. to produce permanent diabetes in cats (Cotes, Reid and Young, 1949) and in dogs (Campbell, Davidson, Snair and Lei, 1950) It would appear that this is the main diabetogenic component of anterior pituitary extracts but the cortical hormone liberated by ACTH and the thyroid product by the thyroid stimulating hormone must also play their part, as described elsewhere Growth hormone stimulates growth in young dogs and does not produce diabetes until the animals have matured (Young) If a part of

the rat's pancreas is removed, but not enough to cause diabetes, this condition may be produced by large doses of growth hormone, thyroxine or cortisone (Houssay). The rat is very resistant to these diabetogenic influences, perhaps in part because of its ability to make more insulin when needed. In the intact rat, however, on a high carbohydrate diet, growth hormone and ACTH (given together) regularly produce hyperglycemia and glucosuria (Engel).

THE GLYCOTROPIC OR ANTI-INSULIN ACTION OF ANTERIOR LOBE EXTRACTS It was first demonstrated by Houssay and Potlick and adequately confirmed by many later investigators that treatment with anterior lobe extract can induce in either normal or hypophysectomized animals an insensitivity to the action of insulin.

Bennett observed hyperglycemia and increased liver glycogen values upon prolonged administration of the adrenotropic hormone. Jensen and Grattan, and Ingle have shown that the adrenotropic substance as well as extracts of the adrenal cortex, and crystalline corticosterone produce a definite glycotropic effect. The adrenocorticotrophic factor failed to produce this anti-insulin effect or to promote deposition of liver glycogen in adrenalectomized mice. These findings, therefore, strongly suggest that the glycotropic effect of anterior pituitary extracts is due to the adrenocorticotrophic factor. A part of the anti-insulin action may be due to the storage of large amounts of glycogen in the liver. This glycogen is presumably available to counteract the hypoglycemic effect of insulin.

It is well established that hypophysectomized animals are unable to preserve their glycogen stores as normal animals do during fasting. It has been found by Russell and Bennett that this function can be restored, under certain conditions, by anterior pituitary extracts. In subsequent experiments it has been shown that this action is not through the adrenal since muscle glycogen can be maintained at a normal level in the absence of both adrenal glands in the fasting animal. In fed animals, however, adrenalcortical activity is required to maintain muscle glycogen. It has recently been shown by Russell and Wilhelm (1950) that purified growth hormone is able to prevent loss of glycogen from skeletal muscle and from diaphragm and to restore heart glycogen to the normal fasting level in hypophysectomized rats. It is unnecessary, therefore, to postulate a separate glycogen-sparing factor.

THE PANCREOTROPIC ACTION OF ANTERIOR

PITUITARY EXTRACTS. In 1933 Anselmino, Herold and Hoffmann reported that frequent injections of anterior lobe extracts in rats produce in a few days an increase in the size and number of the islets of Langerhans. Richardson and Young were not able to confirm the findings under the conditions defined by the German investigators. They were, however, able to show that the daily treatment of rats with crude anterior pituitary lobe extract for a period of two weeks doubled the amount of islet tissue in the pancreas. More recently Marks and Young showed that the insulin content of rat pancreas was greatly increased under these conditions. (See Insulin Content of Pancreas.)

It will thus be apparent that the anterior pituitary gland affects metabolic processes by a variety of mechanisms. The permanent diabetes produced by growth hormone is due to destruction of the β cells of the pancreas, but growth hormone inhibits the peripheral utilization of carbohydrate. It raises the blood sugar and intensifies the diabetes of completely depancreatized animals. It inhibits the "binding" of insulin by muscle and other tissues. ACTH, by liberating adrenal cortical steroids greatly increases gluconeogenesis and probably also inhibits peripheral utilization of glucose. Cortisone under certain conditions also stimulates the islets and thus may modify the diabetic state. The action of growth hormone or cortisone in decreasing peripheral utilization of sugar raises the blood sugar and this is undoubtedly part of the mechanism by which the β cells are stimulated and destroyed.

The anterior pituitary gland also affects carbohydrate metabolism through the thyrotropic and probably also the gonadotropic hormones. Removal of the adrenal cortex, thyroid or gonads produces histological changes in the anterior pituitary indicative of overactivity, i.e. of attempts to stimulate the missing target organs.

Inhibition of the anterior pituitary effect on carbohydrate metabolism by the use of X-rays or by injection of estrogenic substances can be demonstrated in animals and in man but little of practical clinical value can be expected from these procedures.

THE THYROID AND CARBOHYDRATE METABOLISM

The aggravation of diabetes in man by hyperthyroidism and its amelioration by removal of the thyroid establishes a link between this gland and carbohydrate metabolism. It is surprising that

very little influence on the diabetes of depancreatized animals can be demonstrated by thyroidectomy. This may be due in part to incomplete removal of thyroid tissue. When very extensive atrophy of the thyroid is produced by hypophysectomy in dogs, the blood sugar and a normal level of urinary nitrogen can be maintained for long periods during fasting if thyroxin is supplied. Without it hypoglycemia may soon terminate the experiment. Thus it appears that the thyroid may play a role in the effect of the anterior pituitary on carbohydrate metabolism. The slight effect of thyroxin administration on the intensity of pancreatic diabetes in some animals still requires explanation but Houssay, Foglia and their colleagues have shown that in thyroidectomized rats 95% of the pancreas can be removed without causing glucosuria. The administration of thyroid preparations to such animals may result in the development of permanent diabetes (metathyroid diabetes). In some partially depancreatized dogs permanent diabetes could also be induced by large doses of thyroid material.

The administration of thyroid substance or of thyroxin to normal animals has no immediate effect on blood sugar but a loss of liver glycogen may be demonstrated within six hours. There is also apparently a rise in the protein content of liver due perhaps to mobilization from peripheral tissues. Increased gluconeogenesis from protein can be readily demonstrated when thyroid substance is fed. An increase in the d-amino-acid oxidase activity of liver has been reported.

In clinical hyperthyroidism a mild hyperglycemia and glucosuria may be present. A comparable condition may be produced in animals by administration of thyroid material. At this stage liver glycogen is easily mobilizable (the actual amount present may be less than normal) and adrenaline elicits more hyperglycemia and insulin less hypoglycemia than normally.

When thyroid feeding is continued there is a profound decrease in liver, muscle and heart glycogen. In this second stage the animals are resistant to adrenaline and extremely susceptible to insulin. They may exhibit spontaneous hypoglycemia or develop it as a secondary result of a small injection of dextrose. This latter effect may be due to the liberation of insulin the action of which is not buffered by liver glycogen.

Thus the effect of the thyroid on carbohydrate metabolism is the resultant of several actions: (1) increased oxidation of carbohydrate in tissues gen-

erally, (2) the increased rate of hepatic gluconeogenesis, and (3) destruction of islet tissue in partially depancreatized animals.

GLUCAGON

The availability of crude insulin preparations in 1923 enabled Collip, and Kimball and Murlin to detect an initial transient rise in blood sugar in normal animals which received the solutions intravenously. The latter investigators introduced the name "glucagon" which was adopted by Bürger (1929) who, with his colleagues, has been responsible for most of the early studies of the properties of this material. Bürger demonstrated that glucagon could be separated from insulin and that the hyperglycemia was accompanied by a fall in liver glycogen. Adrenalectomy did not eliminate this effect. Shipley and Humel (1945) showed that crude insulin preparations accelerated glycogen breakdown in liver slices. Sutherland and Cori (1948) obtained evidence that this glycogenolysis was due to glucagon and that this substance activated the phosphorylase system. Staub, Sinn and Behrens (1953) have now obtained glucagon in crystalline form as a polypeptide differing significantly from insulin in amino acid content. There are many converging paths of evidence which suggest that this substance is made in the alpha cells of the pancreas and perhaps also in similar cells of the fundic mucosa of the stomach. Sutherland and de Duve obtained glucagon activity from these two sources only. The physiological significance of this hormone is not yet known but it appears to be liberated after growth hormone injection (Bornstein, Reid, and Young 1951, Foa, Magid, and Glassman 1953) and as one of the mechanisms which correct hypoglycemia. This subject has been well reviewed by de Duve (1953) who, with his colleagues Vuylsteke, Cornelis and others, has added greatly to our knowledge. The glycogenolytic activity of glucagon has apparently been demonstrated in human subjects by hepatic catheterization but its clinical significance awaits study.

THE NERVOUS REGULATION OF CARBOHYDRATE METABOLISM

As is the case with many other aspects of carbohydrate metabolism, Claude Bernard paved the way for the investigation of the influence of the nervous system. Bernard (1855) punctured the floor of the fourth ventricle in unanesthetized animals and observed that the piqure produced a

prolonged glucosuria (The blood sugar of a rabbit may rise to 0.40 per cent within an hour and the effect may persist for several days or longer)

It has been appreciated for some time that lesions in the hypothalamic region may cause glucosuria and the relation of these lesions to pituitary secretion has been in doubt. While lesions in this region might interfere with the absorption of the pituitary hormones or more likely, stimulate or destroy the nerve fibers going to the gland, it has been established that stimulation of the thoracic autonomic center in the hypothalamus may produce hyperglycemia. This gave rise to the idea that lesions lower in the brain stem might act by irritating the fiber tracts from the thoracic autonomic center but the situation is apparently not so simple. Donhofier and Macleod attribute special significance to the pons and there may be various centers, the impulses from which affect the level of blood sugar by more or less indirect paths.

In a recent paper, with an excellent review of previous work, Anderson, Rioch and Haymaker (1952) showed that in dogs and rats a prolonged (20-30 days) decrease in glucose tolerance was produced by transection of the brain whether at the pontine, midbrain or hypothalamic level. In animals in which decortication was performed and in those in which the brain stem was hemisected there was no change in glucose tolerance. These findings provide evidence for a homeostatic control of blood sugar level by centers in the brain stem.

It will be appreciated that lesions of the brain involving (1) the pathways which carry glycolytic impulses to the liver or (2) the tracts which supply the adrenals, pancreas or pituitary, may cause serious disturbances of carbohydrate metabolism by interfering with these normal mechanisms. The work of Anderson et al. does not suggest that failure of glycogen formation or increased secretion of adrenal corticoids provides an important part of the answer to this problem. It is apparent that the field demands much further exploration with the light of modern endocrinology and neurophysiology focussed on it.

GLUCOSE TOLERANCE CURVES

When glucose is administered by mouth to a normal animal the blood sugar begins to rise within two or three minutes. This indicates that the sugar solution passes rapidly through the stomach to the duodenum. If large amounts of

sugar are provided there may be considerable loss in the urine, i.e., *alimentary hyperglycemia* and *glucosuria*. After the usual meal, however, the hyperglycemia is not sufficient to produce glucosuria. When moderate amounts of sugar are given the rise in blood sugar is transient and the return to the normal level rapid. This phenomenon is extensively used to test carbohydrate tolerance. The factors which determine the shape of the curve when from 50 to 100 grams of sugar are given by mouth are (1) The rapidity of absorption, (2) the extent of the storage and utilization of glucose by the tissues, and (3) the rate of discharge of sugar from the liver. The first factor may, of course, be eliminated by injecting the sugar intravenously. The reaction of the tissues and of the liver to injected sugar may be direct or indirect. When more sugar is presented to the tissues more is utilized even though the insulin available remains constant. When the blood sugar rises the liver may discharge less sugar (Soskin, Allweiss and Cohn). The hyperglycemia may affect the liver and other tissues indirectly by increasing the insulin output and perhaps by other endocrine adjustments. The insulin liberated increases storage of glucose and fat and may decrease gluconeogenesis. An abnormal curve may indicate inability of the liver or pancreas, or of both to perform their normal functions. It might be due in part or completely to defective oxidation and storage in the muscles. It must be emphasized that an abnormal glucose tolerance curve *does not* necessarily indicate a deficiency of available insulin.

UNDERNUTRITION AND CARBOHYDRATE METABOLISM

In 1873 Lehmann and in 1877 Claude Bernard noted a glucosuria in fasting animals after the administration of carbohydrate, and in 1890, Hofmeister, who made the first quantitative studies, named the condition "hunger diabetes". Utilization of carbohydrates is at a maximum in animals which have been fed on diets rich in these substances. After periods of fasting or of fat feeding, there is a definite impairment of glucose utilization which may easily be detected by the results of a glucose tolerance test. The feeding of an exclusively fat diet produces effects on glucose utilization indistinguishable from those of complete starvation.

¹ Normal dogs may be given 0.85 gram of glucose per kilogram per hour for long periods (Woodyatt) without producing glucosuria. This is approximately the same value as the maximum rate at which glucose is absorbed from the intestinal tract.

Proteins exert an effect intermediate between that of sugar and fat, i.e., some impairment of glucose utilization is produced by an exclusively protein diet. Diets adequate in other respects but providing a low caloric intake cause little disturbance in carbohydrate utilization. The administration to animals or human subjects exhibiting hunger diabetes of a diet containing glucose causes a prompt improvement in carbohydrate tolerance.

The mechanisms of production and alleviation of the defect in carbohydrate utilization produced by fasting are not as yet completely elucidated. It would appear from the recent findings of Chambers, Cori and others that oxidation of glucose in the tissues is interfered with to a much greater extent than is glycogen formation. It has been established that the administration of insulin effects a partial restoration of carbohydrate utilization (Cori and Cori, Dann and Chambers). This finding suggests that insulin liberation may be depressed in hunger diabetes. Himsworth feels that a change in sensitivity to insulin is involved rather than a diminution of pancreatic output. On the other hand, the observation by Haist, Ridout and Best shows that the insulin content of the pancreas of rats may be reduced to nearly half the normal value by starvation or by fat feeding. This finding in conjunction with others supports the view that insulin liberation may be decreased. The evidence, which has been well reviewed by Chambers, does not indicate that the complete explanation of hunger diabetes will be found in the abnormal response of any one organ or tissue. This conception is supported by Ingle's report that insulin alone does not rapidly restore the carbohydrate tolerance of fasted animals to normal and by the finding that at least one important enzyme system (phosphorylase) is increased in activity during starvation (Lundbaek and Goranson). The lipogenic ability of the liver is reduced to the diabetic level by fasting.

GLYCOGEN DISEASE A clinical condition characterized by the enlargement of one or more organs resulting from the accumulation of glycogen has attracted the attention of research workers. The disease usually bears the name of von Gierke who published an autopsy report on a case in 1929. Van Creveld, in the previous year, had concluded that the hepatomegaly which he observed in a young boy was probably due to excessive glycogen deposition. The glycogen deposits may be in the liver, kidney, heart or in other tissues. The disease is characterized by hypoglycemia and ketosis in the fasting condition, by an abnormal effect of adrenaline

which causes only a slight rise in blood sugar and lactic acid but a large increase in ketosis, and by an increased sensitivity to insulin. The glucose tolerance test gives a prolonged hyperglycemia without glucosuria. The glycogen content of the blood is increased and van Creveld noted a resistance of this material to glycogenolysis. The glycogen in liver and kidney does not disappear at a normal rate after removal from the body and an interference with the glycogenolytic process is therefore indicated.

It would appear probable, as van Creveld has suggested, that glycogen disease is a continuation in childhood of a fetal condition in so far as certain aspects of carbohydrate metabolism are concerned. There are large deposits of liver glycogen in the fetus and this material is resistant to the action of adrenaline. It is well established that in some species the fetal pancreas, at or near term, contains very high concentrations of insulin. Hyperinsulinism may produce excessive deposits of glycogen in well-fed animals and ketosis under fasting conditions. It has been suggested that the secretion of the anterior pituitary gland makes liver glycogen more resistant to breakdown. But all these observations must be considered in the light of the findings by G. T. Cori et al. that the structure of glycogen from the liver of these cases is sometimes definitely abnormal and that the abnormalities in response to enzymes differ from case to case. There may, therefore, be many different types of glycogen disease.

ALLOXAN DIABETES This type of experimental diabetes was first produced by Dunn, Sheehan and McLetchie (1943) who showed that alloxan has a selective necrosing action on the Islets of Langerhans. In 1937 Jacobs had noted the effect of alloxan in rabbits—an initial hyperglycemia and a subsequent hypoglycemia. No histological studies were made and Jacobs postulated an insulin-like action of alloxan. It has now been shown that the hypoglycemia is due to liver damage or liberation of excess insulin from damaged islets or to both. The diabetic state is caused by a failure of islet cells to produce insulin. The diabetic action of alloxan has been demonstrated in the rabbit, rat, cat, monkey and dog. An unsuccessful attempt has been made to destroy the abnormally active islet cells in advanced cases of hyperinsulinism in man by the administration of alloxan. Lesions in the liver and kidney, less marked than those produced in the islets, are seen in some species after the injection of this chemical.

There is, as yet, no evidence that alloxan has any physiological significance. A substance resembling it was reported, many years ago, in the urine and intestinal mucus in certain pathological conditions in man.

Griffiths has produced hyperglycemia and glycosuria but not permanent diabetes with uric acid which is chemically related to alloxan. To secure this effect the glutathione level of blood was depleted by a diet deficient in cystine and methionine and feeding large amounts of ascorbic acid. Lazarow in 1945 was the

for to show that glutathione injected just before the diabetogenic dose of alloxan protected the rats. While many other substances such as dehydroascorbic acid, etc., thus protect glutathione as a normal constituent of cells is of particular physiological interest. Glutathione reacts with alloxan to reduce it to dialuric acid which is not diabetogenic. Letch and Bailey found an extensive fall in blood glutathione after alloxan, and previous investigators had shown similar changes in other tissues. The injection of various pituitary preparations has been found to lower the glutathione content of tissues. Cohn found in his studies of transient diabetes in man produced by injecting ACTH that there is a direct correla-

tion between the blood glutathione level and the diabetogenic effect, i.e. the higher the blood glutathione the less the effect. In animals injected glutathione consistently potentiates steroid diabetes by some unknown mechanism. The protective effect against diabetes may be attributed to the removal of diabetogenic compounds from the blood or beta cells. It has been suggested (Hédon, Lazarow) that factors which increase insulin production may exhaust the supply of sulfhydryl groups by appropriating the cysteine for the insulin molecule and thus decrease the available glutathione and in this way make the β cells particularly susceptible to alloxan or naturally-occurring damaging factors.

CHAPTER 50

FAT METABOLISM

CLASSIFICATION OF THE FATS AND FATTY SUBSTANCES OCCURRING IN THE BODY

When a tissue is thoroughly extracted with a mixture of ether and alcohol almost all of the fatty substances are removed

The amount of fatty acids combined in the cerebrosides and phospholipids may be determined by taking advantage of the insolubility of these materials in acetone to which some magnesium chloride has been added. The fatty acids may then be liberated and estimated by one of a variety of procedures. An estimate of the amount of the phosphorus-containing substances can be obtained by determining the phosphorus content of the extract, and similarly with those containing carbohydrate by estimating the sugar. The free cholesterol is removed by precipitation with digitonin and estimated gravimetrically or by colorimetric means. In another sample of the extract the bound cholesterol is liberated by saponification as with sodium ethylate and the total cholesterol determined. The difference between the free and the total is the amount of cholesterol combined with fatty acid and this can be calculated from the known combining weight of fatty acids with cholesterol. The result is an estimate of the amount of cholesterol ester. The difference between the total fatty acid content and the sum of the amounts combined with carbohydrate, phosphoric acid, and cholesterol, gives us an estimate of the amount combined with glycerol. The fatty materials or lipids may therefore be classified as follows

(1) *Fats* Esters of fatty acids and glycerol

(2) *Lipoids*¹

(a) *Phospholipids* (Phosphatides) Fatty substances containing fatty acids, phosphoric acid and some of the following constituents, glycerol, inositol, choline, ethanolamine, serine and sphingosine. Examples are—lecithin, cephalin and sphingomyelin

(b) *Cerebrosides* (Glycolipids) Combinations of fatty acid, sugar and sphingosine (phrenosin, kersin, etc.)

¹ The phospholipids, cerebrosides and waxes which resemble fats may be termed lipoids or fat-like substances. The sterols and hydrocarbons while associated with fats are chemically quite distinct.

(c) *Waxes* Esters of fatty acids and certain alcohols (but *not* glycerol), (cholesterol esters, beeswax, etc.)

(3) *Sterols* Hydrogenated phenanthrene derivatives (free cholesterol, ergosterol, etc.)

(4) *Hydrocarbons* (Squalene, carotene, etc.)

THE TRIGLYCERIDES

CHEMICAL STRUCTURE In the members of the first group, the fats, one molecule of glycerol is combined with three of fatty acid. Oleic acid $C_{18}H_{34}O_2$, stearic acid $C_{18}H_{36}O_2$, and palmitic acid $C_{16}H_{32}O_2$, are the three which account for the bulk of the fatty acids of the neutral fat fraction of body tissues. The latter two are saturated, while oleic acid has a double bond in the middle of its fatty acid chain. This renders it less stable and more easily oxidized. Oleic acid is liquid at low temperatures while stearic and palmitic are solid even at body temperature. It is now thought that the glycerides are usually mixed, i.e., they contain two or more different fatty acids in their molecule. Fatty acids more unsaturated than oleic (linoleic and arachidonic) are also found in the triglycerides.

The neutral fats obtained from animal tissues contain fatty acids which have an even number of carbon atoms. The characteristic fatty acids in the depot fat vary with different animals. Hilditch and Lovern have shown that certain fatty acid mixtures are characteristic of marine animals, others of fresh water fish and others of land animals. The fat of the fish, for example, contains large quantities of higher unsaturated fatty acids C_{20} and C_{22} , but in the depots of ox and pig these are almost absent.

DISTRIBUTION Reserve fat is found in the so-called interstitial tissue of all organs with the exception of the brain. There may be very large masses of this material in subcutaneous tissue and in various other places such as in the omentum, the perirenal fat depot and so on (see table 65).

The distribution of fat has been found to be independent of the type of diet which, however, controls the amount and in many species the character of the fat deposited.

PHYSIOLOGY The main function of the glycerides is undoubtedly to provide a source of energy.

TABLE 65

Distribution of tissue lipids

Column (1) (after Campbell and Lucas) gives the total lipids of the tissues of mice as per cent of body weight.

Column (2) shows the distribution of adipose tissue in the rat (after Mendel) as per cent of total adipose tissue.

	(1)		(2)
Carcass	34	Subcutaneous	50
Adipose tissues	32	Genital	20
Skin	18	Perirenal	12
Liver	5	Mesenteric	10
Intestine and spleen	3	Intermuscular	5
Genital organs	3	Omental	3
Kidneys and adrenals	2		
Brain	2		
Lungs and thymus	1		

Fat not only furnishes more energy per gram (9.3 calories) than carbohydrate and protein but it is the only food material stored in the dry state. The total fat content of a well-nourished animal may be between 10 and 12 per cent and under unusual conditions much higher. A very large proportion of the total energy store of the body is in the form of fat.

While animals grow normally when diets containing only very small amounts of fat are provided, several investigators have reported that growth is seriously retarded on a diet free of fat (McAtee, Anderson and Mendel). Palmitic, stearic and oleic acids are not essential but Burr and Burr have provided evidence that certain of the unsaturated fatty acids are necessary. Growth is favorably affected by linoleic, linolenic and arachidonic acids and the development of skin lesions (similar to acrodermia) is prevented. These acids should be considered separately in dietary studies since there are definite differences in the magnitude of their effects on growth and on the skin lesions. There is an interesting interrelationship of the actions of the unsaturated fatty acids and those of pyridoxin, pantothenic acid and other accessory food factors.

In some species, of course, the layers of fat in the subcutaneous spaces serve as an insulating mechanism against extreme cold. Thus in cold climates the fat deposited just under the skin contains a relatively large amount of oleic acid. It is therefore more liquid than the material which is

found in the less superficial subcutaneous reserves. Henriques and Hanson have pointed out the fact that since the temperature of the outermost subcutaneous fat is appreciably lower than that of the deeper parts, the body requires less congealable type of fat at this place. In a somewhat similar investigation of fat in different parts of the body, Anderson and Mendel could find no difference in saturation of the fat from the superficial subcutaneous region and that of deeper and warmer parts of the body. These latter workers used animals which are not normally subjected to extremely low temperatures.

Recovery from the depot fat of a particular dietary fatty acid has been repeatedly demonstrated. This deposition is, however, limited to the higher fatty acids (above C_{12}). Interconversion of the fatty acids within the body is well established. Schoenheimer and Rittenberg have shown that the length of the carbon chain of fatty acids can be increased or decreased two carbon atoms at a time and further that desaturation or hydrogenation of unsaturated acids can take place in the body. While one double bond can be formed the body appears unable to introduce more than one. Thus the highly unsaturated fatty acids such as linoleic and linolenic are not found in fat which has been synthesized in the body. They must be provided in the food.

Conversion of carbohydrate and protein to fat

The formation of body fat from carbohydrate of the diet was established by the classical work of Liebig (1852) and Lawes and Gilbert (1853). The rate at which the fatty acids are formed has now been studied by the use of isotopes. These studies have shown that conversion of carbohydrate to fat proceeds continuously and does not occur only (as was formerly believed) when there is excessive intake of carbohydrate.

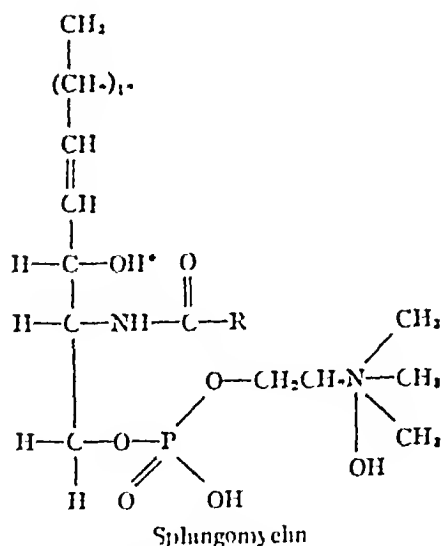
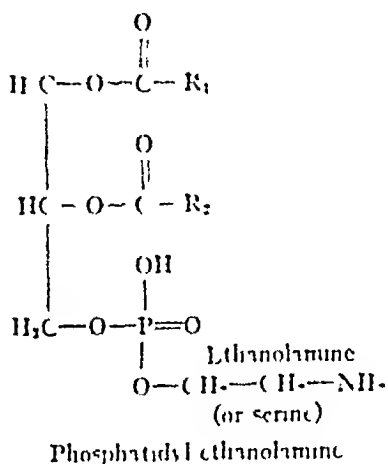
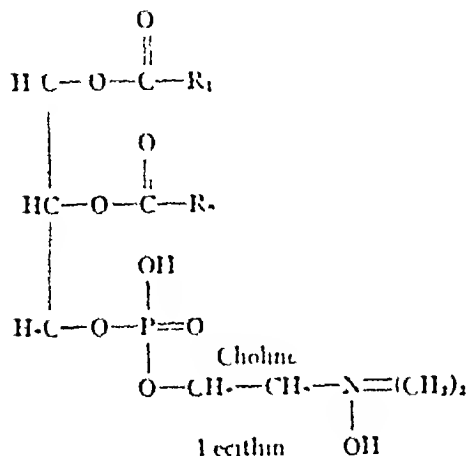
The formation of carbohydrate from protein is well established. It has been assumed that this newly formed carbohydrate is available for fat synthesis and good evidence for this assumption has been secured by Longenecker and by Hoagland and Snider. On an almost exclusively protein diet the fat formed was similar in composition to that synthesized from dietary carbohydrate.

Many of these recent additions to our knowledge establish the fact that the fat depots are not essentially inert storehouses of energy as was previously thought, but are centers of continuous metabolic

activity. Glycogen and fat formation can take place in adipose tissue.

THE PHOSPHOLIPIDS

CHEMICAL STRUCTURE The phospholipids probably occur mainly in combination with tissue



* May be esterified with fatty acid

proteins. They are divided into the monoaminophospholipids in which the ratio of nitrogen to phosphorus is 1:1 and the diaminophospholipids in which the ratio of nitrogen to phosphorus is 2:1. The identified monoaminophospholipids are lecithin (phosphatidyl choline), phosphatidylethanolamine and phosphatidylserine. These two latter compounds with some inositol containing substances and uncharacterized basic materials have been referred to as the "cephalin" complex.

Lecithin and the members of the cephalin complex are composed of glycerol and fatty acids as are the triglycerides, but one fatty acid may be considered to be replaced by the phosphoric acid-nitrogenous base complex. There may be one saturated and one unsaturated fatty acid but in some cases two saturated or two unsaturated fatty acids occur. Oleic and either palmitic or stearic acids are commonly found but more highly unsaturated fatty acids such as linoleic and linolenic have been identified. In recent years pure synthetic lecithins identical in every respect with the natural substances have been made by Baer and his associates. Their work indicates that the naturally-occurring lecithins are of the alpha type. Sphingomyelin is a diaminophospholipid and contains two bases, choline and sphingosine, and one fatty acid radical but no glycerol. Different sphingomyelins containing respectively, stearic, lignoceric and nervonic acids have been found.

DISTRIBUTION The phospholipids are widely distributed in the body and it is thought that all cells contain one or more of these compounds. Sphingomyelin is present in much larger amounts in the brain and nerve tissues than elsewhere and is usually associated with the cerebroside.

The "cephalin" content of the brain is considerably higher than the lecithin concentration. In the liver and spleen they occur in almost equal amounts while in the kidney, heart and lung, lecithin predominates in a proportion of about two to one. "Cephalin" and lecithin have been isolated from the gastric mucosa of pigs but sphingomyelin could not be demonstrated.

The evidence of the French workers Mayer and Schaeffer, and Terroine showed that the amount of phospholipid in a particular organ in a given species remains relatively constant under a variety of conditions including extreme starvation. As a result, they have called this fraction the "élément constant". This evidence strongly suggests that the lipids are structural components of the cell. Their rapid rate of turnover like that of many

TABLE 66

*The lecithin, cephalin, and sphingomyelin content of normal human organs**

ORGAN	LECITHIN†	CEPHALIN†	SPHINGOMYELIN†
Brain‡	4 81	20 42	5 66
Lung	3 85	2 00	1 45
Spleen	3 54	4 16	0 86
Kidney	5 10	3 26	0 72
Liver	4 81	4 62	0 38
Heart	4 47	2 06	0 34

* From Thannhauser et al (J Biol Chem, 129, 717, 1939)

† The values representing mg per 100 mg of dried organ

‡ Including both white and grey matter

other structural elements has been established. In table 66 the phospholipid content of various human tissues is given.

PHYSIOLOGY The exact role of the phospholipids is not yet understood but their metabolic importance is widely accepted. As noted above the phospholipid content of the various tissues is much more constant than that of neutral fat. Significant variations in phospholipids may, however, be induced by changing the nature of the diet or by alterations in the hormonal balance.

Insulin decreases the plasma phospholipid. Repeated injections of thyroid substance increase the amount in liver while removal of the thyroid causes a decrease. With the onset of pregnancy in the human subject the plasma phospholipids show a rise which continues to term (the other lipids also increase). The work of Aten and Hevesy indicates that milk phospholipids are synthesized in the mammary gland.

Bloor has suggested that the phospholipid content of a tissue may be considered an index of the extent and variety of its physiological functions. Thus, the phospholipid content is increased with physiological activity and decreased when the cells become less active. A secreting salivary gland of the dog (Caminade, Mayer and Vallée) has a higher phospholipid content than the resting one. On the other side the development of the corpus luteum is accompanied by a very significant increase in its phospholipid content (Bloor, Okey and Corner). When the body temperature of rabbits and dogs was greatly reduced by immersion in very cold water, Mayer and Schaeffer found a decrease in the phospholipid content of the liver and a compensatory rise when the animal had re-

covered from the experience. The phospholipid content of rapidly growing malignant tumor cells is higher than in normal tissue or in benign tumors of the same tissue.

The metabolic function of the phospholipids is most readily studied experimentally. The injection or feeding of synthetic compounds containing labelled atoms such as deuterium or radioactive tritium in place of hydrogen, heavy carbon C^{12} , radioactive carbon C^{14} , heavy nitrogen N^{15} , and radioactive phosphorus P^{32} has made it possible to study the synthesis of the phospholipids. Schoenheimer, Bloch, Rittenberg, Hevesy, Artom, Chaikoff, Stetten and many others have used these techniques to increase our knowledge of phospholipid metabolism. Sinclair used elaidic acid, the unnatural trans isomer of oleic acid, as a labelled compound in similar studies. The assumption that the body is unable to distinguish between the cis- and trans-isomer may not be justified. Indeed, recent reports of studies with compounds doubly or trebly labelled have shown that sometimes the different tracers have given different answers. Caution is necessary in the interpretation of data obtained with isotopes.

These methods have shown that the turnover of phospholipids is most rapid in the mucosa of the small intestine during fat absorption, somewhat slower in the liver and much slower in the muscles and kidneys. In the brain the incorporation into phospholipids of P^{32} (given as inorganic phosphate) is a slow process. The correspondingly slow loss of P^{32} from the brain confirms the sluggish turnover of phospholipids in nervous tissue. The limited data on the turnover rates of sphingomyelins in liver and muscle indicate that they are of about the same order of magnitude as that of lecithin although the latter always has the fastest regeneration rate.

A close relationship between liver and plasma in regard to phospholipids has been established. Chaikoff and his associates have shown that virtually all the plasma phospholipids are derived from the liver. This was demonstrated by administering phosphate labelled with P^{32} and palmitic acid labelled with C^{14} to normal and hepatectomized dogs. Only traces of isotope were recovered from the plasma of hepatectomized dogs although other tissues (kidney, intestine, etc.) were able to synthesize phospholipids at the same rate as in the control dogs. The liver appears also to be the tissue principally concerned in the utilization of plasma phospholipids since their concentration

changes but slowly in the absence of the liver. The incorporation of administered P^{32} into phospholipids of the corpuscles is a much slower process than into those of the plasma.

THE CEREBROSIDES (GLYCOLIPIDS)

In cerebroside the base sphingosine is combined with a sugar (galactose or glucose) and a fatty acid. Thus the cerebroside differ from one another in the nature of their fatty acids or carbohydrate. They can be distinguished from the phospholipids by the absence of phosphorus. The only individual cerebroside that have been isolated in a pure condition contain galactose, they differ in their fatty acid component. Sphingosine contains a characteristic saturated acid (lignoceric, $C_{24}H_{48}O_2$), phrenosin (also called cerebrin) contains the corresponding α hydroxy acid (cerebronic, $C_{24}H_{46}O_3$), nervone contains the corresponding Δ_{15} unsaturated acid (nervonic, $C_{24}H_{46}O_2$), hydroxynervone contains the corresponding unsaturated α hydroxy acid (hydroxynervonic, $C_{24}H_{46}O_3$). They are found particularly in nervous tissue, only minute amounts occur elsewhere in the normal body.

CHOLESTEROL

Chemical nature Cholesterol is found associated with the fats but chemically it is not related to them. Cholesterol, a white waxy solid, is the principal sterol found in animal organisms. The sterols are crystalline saturated or unsaturated complex alcohols derived from cyclopentanoperhydrophenanthrene. They occur both free and combined as esters or glycosides. Very minor changes in their structure sometimes produce extraordinary differences in physiological activity. The term steroid is applied to a group of closely related compounds containing the same complex nucleus. These substances include the sex hormones, bile acids, cardiac glycosides, toad poisons and vitamin D. Cholesterol has been known since the eighteenth century as the chief component of human gall stones. It is an unsaturated secondary alcohol. Most of the cholesterol occurs in the body as the so-called free form (the unesterified alcohol) which is precipitable by digitonin. A much smaller portion is present as the ester of long chain fatty acids (the so-called bound cholesterol). Recent studies indicate that both forms of cholesterol are present in some loose combination with protein. These complex materials of very high molecular weight, referred to as lipoproteins, have been studied by Chargaff, Gofman and others.

TABLE 67
Total cholesterol content of normal tissues
(per cent of fresh tissue)

	RABBIT*	MAN†
Adrenals	7.3	4.74
Brain	1.8	1.93
Skin		0.93
Kidney	0.44	0.33
Spleen	0.38	0.36
Lung	0.38	
Liver	0.29	0.32
Adipose tissue (subcut)		0.24
Hair		0.17
Blood (whole)	0.08	0.17
Plasma	0.06	0.23
Corpuscles	0.12	0.12
Heart		0.14
Muscle	0.06	0.07

The tissues of a man weighing 70 kg may be calculated to contain slightly over 100 g of total cholesterol: about 30 g in adipose tissue, 25 g in brain, 20 g in muscle, slightly over 5 g in whole blood, 5 g in liver, because of their small size adrenals account for only about 0.5 g.

* After Chamberlain

† After Cook

DISTRIBUTION Cholesterol is found in all cells and fluids of the body. The free and bound forms are not equally distributed. The free cholesterol content of any particular tissue is characteristic and normally remains relatively constant, the esters, on the contrary, vary considerably in amount with changes in dietary, hormonal and other factors. The brain and suprarenals have the richest supply of cholesterol. In the former it is found mainly in the free form and in the latter about 90 per cent as ester. In bile it occurs only in the free state. In the corpuscles of human blood cholesterol exists chiefly in the free state while in the plasma more than half is present as the ester. In most other tissues the ester makes up a small fraction of the total. Table 67 shows the average total cholesterol content of normal tissues of rabbit and man.

ABSORPTION, TRANSPORT AND EXCRETION OF CHOLESTEROL Crystalline cholesterol administered orally is absorbed in only small amounts unless some fatty material is also present in the intestine. Bile and pancreatic juice are said to aid in its absorption. Colloidal or amorphous cholesterol may be absorbed in the absence of dietary fat.

Combination with the bile acids increases the solubility of cholesterol in intestinal fluids The experiments of Mueller in 1915 on the absorption of cholesterol have been confirmed and extended by the recent studies of Chaikoff and his associates using cholesterol labelled with C^{14} at carbon atom 4. When free cholesterol is fed a portion is esterified and about half appears in the chyle in the bound form, on the other hand when cholesteryl esters are fed some hydrolysis occurs and again about half appears in the lymph in the free form. Thus, the processes of esterification or of hydrolysis (either before, during or after absorption) lead to essentially the same proportions of free to bound cholesterol in the lymph. The exact site of this esterification whether in the lumen or intestinal wall is at present undecided. The increased cholesterol of the blood during fat absorption is probably due in large part to absorption of cholesterol contained in the pancreatic and intestinal juices and the bile. The increase in cholesterol ester has been taken to indicate that the sterol may play a part in the transport of fatty acids. It is very interesting that closely related sterols (phytosterol, coprosterol) are absorbed only in very small amounts (Schoenheimer and Sperry). The blood contains an enzyme which will split the esters and presumably, under other conditions, will cause synthesis (Shope, Sperry and Schoenheimer). Part of the absorbed free cholesterol, but not the ester, is excreted in the bile and part is changed to coprosterol by hydrogenation and eliminated in the feces. Some unchanged cholesterol is also excreted. The cholesterol plus coprosterol in the feces is usually greater than the cholesterol content of the diet. Some of the biliary cholesterol is reabsorbed in the small intestine. Gardner and Gainsborough state that cholesterol can be found in normal urine and that the amount is increased by cholesterol feeding and in certain diseases. Under these circumstances cholesterol deposits may be found in the kidney tubules.

THE PHYSIOLOGY OF CHOLESTEROL Synthesis and degradation of cholesterol go on simultaneously in the animal body. Herbivorous animals do not get any cholesterol in their food yet their tissues contain essentially the same levels of cholesterol as do those of omnivores and carnivores. Synthesis of cholesterol has been shown to occur in many tissues (adrenal, kidney, testis, ovary, small intestine, skin, etc.) but the liver is the most important site. Plasma cholesterol comes chiefly from the liver. The "two-carbon fragment", now

identified as acetyl coenzyme A, has been recognized for some years as a metabolic intermediate in the synthesis of cholesterol in the body. Attempts have been made to measure the rate of synthesis of cholesterol by feeding acetate labelled with deuterium or C^{14} . The time required to synthesize an amount of cholesterol equal to the amount present in the body is called the regeneration time. Schoenheimer and Rittenberg suggested the term "regeneration" to denote replacement by synthesized molecules, since "turnover" implies only replacement without regard to source. Turnover could, for example, be applied to the replacement of tissue cholesterol by labelled dietary cholesterol. The regeneration time can be shown mathematically to be $1.44 \times$ the "half-life regeneration time" as determined by standard techniques.

Rittenberg and Schoenheimer found the regeneration time for total body cholesterol in the mouse to be about 25 to 30 days. The regeneration time of cholesterol in the liver has been estimated to be about 4 to 5 days in the rabbit (Popjak) and in the rat about 9 days (Bloch). Regeneration in the carcass is very much slower, about 48 days being required in the rat. The regeneration time for plasma cholesterol has been found to be about 12 days in both man and dog. The rate of synthesis of plasma cholesterol in normal man appears to be about 500 to 600 mg per day (Rittenberg). The nature and caloric value of the diet affect the rate of synthesis of cholesterol, during fasting the process is much retarded. The amount synthesized per day in man is probably about ten times the average daily intake. An increase in the dietary intake decreases the amount formed in the body.

Hormonal factors also affect the metabolism of cholesterol. A single injection of an anterior pituitary extract containing the adrenocorticotrophic hormone (ACTH) causes a definite fall in adrenal cholesterol within a few hours (Sayers, Long). When the thyroid is overactive, the concentration of cholesterol in the blood is decreased, drugs such as thiouracil, which depress the activity of the thyroid, increase the level of blood cholesterol. In human subjects estrogens depress serum cholesterol. Cyclic variations of serum cholesterol have been described in women during both the menses and pregnancy. On the other hand, there is no convincing evidence for a primary effect of androgens or progesterone upon cholesterol metabolism.

Bloch has demonstrated with the aid of deuterium the conversion of cholesterol to pregnane-

precursor of other steroid hormones

Ever since Overton's work in 1899, the lipid nature of the cell membrane has aroused interest. The hydrophilic nature of the lecithins, the hydrophobic nature of cholesterol and the peculiar intermediate properties of the cholesteryl esters have been invoked to explain the phenomena at membranes. While there are many examples of the physicochemical antagonism of cholesterol and lecithin, little exact information is available as to the physiological role of the cholesterol moiety of the lipoproteins in cell membranes.

THE ABSORPTION OF FAT

(See also p. 541)

Since the fats and their hydrolytic products, the fatty acids, are very sparingly soluble in water, the explanation of their absorption presents special difficulties. The exact mechanism is still unsolved. There are two main schools of thought concerning the processes involved. One group champions Pflüger's lipolytic theory, the other Frazer's partition hypothesis. Nearly 200 years ago Hewson (1774) observed that the milky fluid seen in the lacteals following a meal leaves a grease spot when dried on paper. Under the microscope the milky fluid was seen to contain numerous tiny particles for which Gage in 1920 proposed the name chylomicron. In 1856 Claude Bernard had demonstrated that a fine emulsification of fat occurs in the intestine and showed that this depends upon the presence of bile and pancreatic lipase. Bernard also noted that lipemia was present in the systemic circulation following ingestion of a fat-rich meal. Further study of this phenomenon was made by Munk and Rosenstein in 1891, utilizing a patient with a lymphatic fistula of the leg.

The mode of transfer of the neutral fat from the epithelial cells to the central lacteals remains a matter for conjecture. The leukocytes have been credited with this function but most of the evidence is against their playing such a rôle. During fat absorption the central lacteals show rhythmical contractions which evidently serve to pump the chyle contained within them into the lymphatics tributary to the thoracic duct. Thus the absorbed fat is constantly propelled along the lymphatic channels of the mesentery.

According to the classical lipolytic theory, fat is emulsified and split completely to fatty acids and

glycerol. The fatty acids are rendered water-soluble by complex formation with bile salts, are believed to enter the epithelial cells of the intestinal villi where they are combined with glycerophosphoric acid or phosphatidic acid to form phospholipids and finally reappear as neutral fat. This is carried largely in the lymph, by way of the lacteals, to the thoracic duct where it enters the systemic circulation.

The partition theory suggests that it is not necessary for all the fats to be completely hydrolyzed. Frazer believes that free fatty acids are absorbed by the blood and pass to the liver by way of the portal vein, while glycerides (tri-, di- or mono-), with particle size below 0.5 microns are absorbed through the intestinal mucosa directly into the lymph. Thus absorption of fat is partitioned between portal blood and lymph according to the degree of hydrolysis.

Recently studies have been made with free fatty acids and triglycerides labelled with C^{14} . When these compounds were given by stomach tube to rats, over 90 per cent of the absorbed fatty acids were recovered from the thoracic duct lymph in both cases. Such findings are *not* consistent with the original partition theory. The main tenet of this theory, i.e. that complete hydrolysis of glycerides is not necessary, is however now well established. By feeding synthetic glycerides labelled with C^{14} in the glycerol part of the molecule, it has been conclusively shown that hydrolysis of long chain glycerides in the intestine is neither rapid nor complete. Other studies indicate that in some cases a large proportion of dietary fat is absorbed in the form of glyceride. Small amounts of monoglyceride, formed in the intestinal lumen by the hydrolytic action of lipases, can exert a powerful emulsifying effect, thus enabling the bulk of the dietary glycerides to be dispersed to particles of extremely fine size. These particles, as noted above, can be directly absorbed.

Fat is not only transported as chylomicrons which are largely triglycerides but as phospholipids and cholesterol esters, and in lipoproteins. The relative importance of these vehicles is a subject for further research.

THE ELIMINATION OF FAT FROM THE BODY

There is no fat or phospholipid in the urine under normal conditions but the latter may appear in disease. There are only traces of fat in the secretion of the skin.

TABLE 68

Lipid components of plasma in the fasting state (mean values in mg per 100 ml) After Boyd

	MAN	RAT	RABBIT	COCK PHEAL
Total lipid	530	230	243	520
Total fatty acid	316	152	169	361
Neutral fat	142	85	105	225
Phospholipid	165	83	78	155
Total cholesterol	152	52	45	100
Free cholesterol	46	21	22	34
Cholesteryl esters*	178	52	39	111

* Bound cholesterol $\times 1.68$

The fat content of the feces is normally between 6 and 12 per cent of the fat which is absorbed but as Sperry has pointed out this bears no relation to the food fat. In dogs, even after five weeks or more on a fat-free diet, there are considerable amounts of fat in the feces. Some 60 per cent of this is neutral fat and 40 per cent fatty acids. A large part of this lipid excretion is derived from bacteria and cellular debris.

When the bile duct is ligated, large, light colored stools are usually observed as this procedure interferes greatly with fat absorption. Digestion is nearly complete, however, as is shown by the fact that the fat is excreted chiefly as fatty acids. Fatty stools are found also when the pancreatic ducts are tied although the absorption of fat is by no means completely prevented. When pancreatic

lipase is excluded in this way the excreted fat contains more glyceride and less fatty acid.

BLOOD FAT

Blood transports not only the fatty materials absorbed from the intestines but also lipids resulting from intermediary metabolism. The amount of lipid material found in the plasma of animals in the postabsorptive state (i.e. at least 12 hours after the last meal) does not vary greatly under normal conditions. Table 68 gives typical values for the lipid components in the plasma of several species. If, however, a sample of blood is withdrawn 3 to 6 hours after the ingestion of a large amount of fat, such as olive oil or thick cream, a distinct milkiness of the plasma will usually be observed. The milkiness is due to the presence of the tiny microscopic fat particles called chylomicrons (fig. 50.1). This is a physiological occurrence which is often associated with the postprandial hyperlipemia following a fatty meal. The extent to which the amount of fat in the blood has increased is sometimes studied by counting these particles under a dark field microscope. Erroneous impressions may be obtained by this method, however, since an increased amount of heparin in the blood, and possibly other factors, can minimize or prevent the appearance of this milkiness in plasma which, by chemical methods, can be demonstrated to contain abnormally large amounts of fatty material. Thus, great variations in the degree of milkiness may be observed in a single individual under apparently identical experimental conditions. Part of the discrepancy can be explained in other ways, since the amount of fat present in the blood at any given instant depends on the balance between the rate of absorption and the rate at which it is being utilized or deposited in adipose tissue. It is obvious, therefore, that the degree of alimentary lipemia is not a consistently reliable index of the rate at which fat is being absorbed. The postprandial increase in blood lipids is due chiefly to neutral fat, smaller increases in cholesteryl esters and phospholipid are often observed. Besides the transient alimentary hyperlipemia there are the so-called retention and transportation hyperlipemias. As stated above, the cholesterol and phospholipids found in plasma originate chiefly in the liver. The plasma phospholipids consist chiefly of the choline containing types (lecithin and sphingomyelin). The larger proportion of the lipid components of the plasma appear to be in combination with proteins.

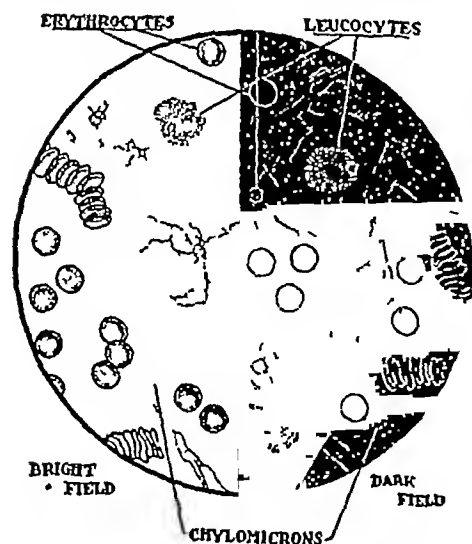


FIG. 50.1 Shows chylomicrons (After Gage.)

FAT TRANSPORT

The long held belief that phospholipids are particularly active in fat metabolism and transport, seems now to be questioned. Pihl and Bloch, for example, using acetate labelled with C^{14} , have shown that the rate of regeneration of neutral fat is faster than that of phospholipid, from which they concluded that phospholipids are not obligatory intermediates in the synthesis of neutral fat. Their data on plasma lipids suggest that it is unlikely that phospholipids are a major vehicle for fat transport. Indeed, the data suggest that neutral fat may represent the most important means of lipid transport in the plasma.

A relatively unexplored possibility is that the caloric-rich lipids in blood may be used to supply energy to areas of higher metabolic activity. To explain the large amounts of lipids found in plasma was not easy in view of the well-known insolubility of neutral fats and free cholesterol in water. It was first suggested that incorporation of fatty acids into phospholipids and cholesteryl esters would account for the findings. The more recent discovery that these latter substances and neutral fats occur in the plasma largely in combination with protein, may provide a hypothesis to account for the transport of energy-rich compounds in the form of lipids. It is conceivable that the lipoproteins may transport water-soluble lipids in the plasma by some mechanism similar to that in which hemoglobin transports oxygen (Turner).

THE LIVER AND FAT METABOLISM

Recent tracer studies have confirmed the importance of the liver in the metabolism of fat. As mentioned earlier, the liver is the main site of formation of plasma phospholipids and of plasma cholesterol. The liver is the major organ for the synthesis of fatty acids from carbohydrate (although this process has been shown to occur in other tissues, e.g. adipose tissue, kidney, muscle, gastro-intestinal mucosa and lung) and it is the principal site for the oxidation of the higher fatty acids. Some of the earlier observations which linked the liver with fat metabolism are as follows: (1) The fatty acids combined as liver phospholipids and glycerides are more unsaturated than the fatty acids found in other tissues. (2) The rate of phospholipid turnover in the liver is more rapid than in any other tissue, with the possible exception of the intestinal mucosa during fat absorption. (3) The liver is the principal site of formation of the

ketone bodies which are well-known metabolic intermediates in the metabolism of fat.

(1) The fact that the fatty acids in the liver are more unsaturated than those in other tissues was interpreted by Leathes and Meyer-Wedell as evidence for desaturation of fatty acids (formation of double bonds) in the liver. Raper suggested that the liver cells may selectively retain the more unsaturated fatty acids. Using fatty acids labelled with deuterium Schoenheimer and Rittenberg were able to supply direct evidence for the desaturation of stearic acid to oleic acid; they also showed that degradation of stearic acid to palmitic acid (with two less carbon atoms) occurs in the body. The lipids of the body were shown to be constantly undergoing synthesis, interconversion and degradation and the rate at which these processes occur was found to be much greater in the liver than in other organs. More recent studies using fatty acids and glycerides labelled with C^{14} have emphasized the important role of the liver in the utilization of fat. Experiments on hepatectomized and eviscerated animals have revealed the possible importance of lipids as sources of energy in extra-hepatic tissue.

(2) The rate of phospholipid turnover in the liver has been studied by giving radioactive inorganic phosphate (P^{32}), fatty acids labelled with C^{14} and choline labelled with heavy nitrogen (N^{15}). All these experiments have shown that the liver phospholipids are regenerated rapidly. More recent studies indicate that the several parts of the molecule (choline, P and fatty acids) are turning over at about the same rate.

(3) In the normal animal the first stage in the oxidation of fatty acids, i.e. conversion to ketone bodies, takes place in the liver. It has been well-established that this organ is by far the most important site for the production of ketones (see below).

The liver normally contains about 6 per cent of total lipids. Variations within the range from 5 to 7 per cent are considered normal, about 50 per cent of this lipid material consists of phospholipids, about 3 per cent of free cholesterol, about 1 per cent of cholesteryl esters and the remainder (45 to 55 per cent) consists principally of glycerides.

The amount of fat which is present in the liver at any time depends on the following factors:

- 1 The rate at which fat is brought to the liver by the blood
- 2 The rate at which the liver is able to take up fat from the blood

- 3 The rate at which the liver can deal with fat,
 - (a) by direct oxidation in the liver cells,
 - (b) by passing it on in the same or in a slightly changed form to other parts of the body
- 4 The rate of synthesis of fat in the liver from other materials

There is an increase in the fat of the liver within a few hours after the ingestion of a meal containing large amounts of fat. Under these conditions the ingested fatty acids appear in the liver lipids thus altering their composition. Under certain conditions the liver may become intensely fatty when no fat is fed, here the fatty acids are derived from the depot fat or from ingested carbohydrate or protein. When "ear-marked" fats are deposited in the tissues and then some poison such as phosphorus or chloroform is administered the excess of liver fat contains the fatty acids which were present in the depot fat.

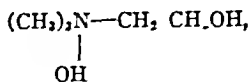
FATTY LIVERS Under a variety of conditions neutral fat or cholesteryl esters may accumulate in the liver to a very much greater extent than in other tissues. In experimental animals fatty livers have been produced containing over 50 per cent of total lipids but this is an extreme situation and values between 20 and 40 per cent are more frequently encountered. These fatty livers are much larger and paler than normal livers, the greater part of the increase in weight of the liver is due to accumulation of glycerides and water. Phospholipids and free cholesterol remain remarkably constant, cholesteryl esters increase significantly whenever glycerides increase.

Some conditions which result in an excessive deposition of fat in the liver are as follows

- 1 Poisons—phosphorus, chloroform and other chlorinated compounds, benzol, ketene, phloridzin, etc.
- 2 Infectious processes
- 3 Environmental changes—reduced atmospheric pressure (e.g. at high altitudes) or elevated temperatures
- 4 Injection of a fraction obtained from the anterior pituitary gland—both purified growth hormone and ACTH are active and the possibility remains that a more specific unidentified factor is operating
- 5 Dietary—diets deficient in choline or its precursors, diets rich in fat or containing excessive amounts of cholesterol, starvation
- 6 Clinical conditions in which one or more of the above situations are present, e.g.

kwashiorkor (world-wide malignant under-nutrition), diabetes, pernicious anemia, yellow fever, pregnancy, etc.

THE LIPOTROPIC FACTORS Choline and its dietary precursors, betaine and methionine, are necessary in the rations of animals in order to prevent the accumulation of excessive amounts of fat in the liver. Choline, a quaternary ammonium base



occurs widely distributed in nature although it is rarely found in the free condition in more than traces. It was first isolated (from bile) by Strecker in 1849 and has long been known as a constituent of lecithin and sphingomyelin. Lean meats (muscle) contain about 100 mg choline per 100 g of fresh meat. Kidney contains 200 to 300 mg, brain 350 to 450 mg, liver from 450 to 600 mg, fish muscle 50 to 80 mg, whole egg 350 to 700 mg, egg yolk about 1400 to 1700 mg, whole wheat 50 to 100 mg, wheat germ 350 to 400 mg, oats, barley, rye, corn and rice contain about 60 to 100 mg, yeast (brewer's dried) 240 to 360 mg per 100 g. Vegetables as a group are low in choline and fruits are very low. Edible fats as obtained commercially are almost free from choline.

The other naturally-occurring substances which possess lipotropic activity are betaine (the internally neutralized acid formed by oxidizing the primary alcohol group of choline), methionine and β propiothetin. They have been shown to possess labile methyl groups and it is believed that they possess lipotropic activity by acting as precursors of choline. Vitamin B₁₂ and folic acid, which are important in hematopoiesis, are also able under some conditions to produce a lipotropic effect. The relations between these two substances in lipotropic phenomena are not yet clearly understood but vitamin B₁₂ appears to be concerned in the biosynthesis of labile methyl groups and folic acid in transmethylation reactions or in the synthesis of some other part of the choline molecule. Inositol exerts a limited lipotropic effect in fat-free diets but this effect is minimized when any fat is in the food. The crude pancreatic extract which has been referred to as "lipocaine" was at one time thought to possess some unique lipotropic influence since it was believed to be free from choline. While it may contain only traces of free choline its content of bound choline sometimes makes it one of

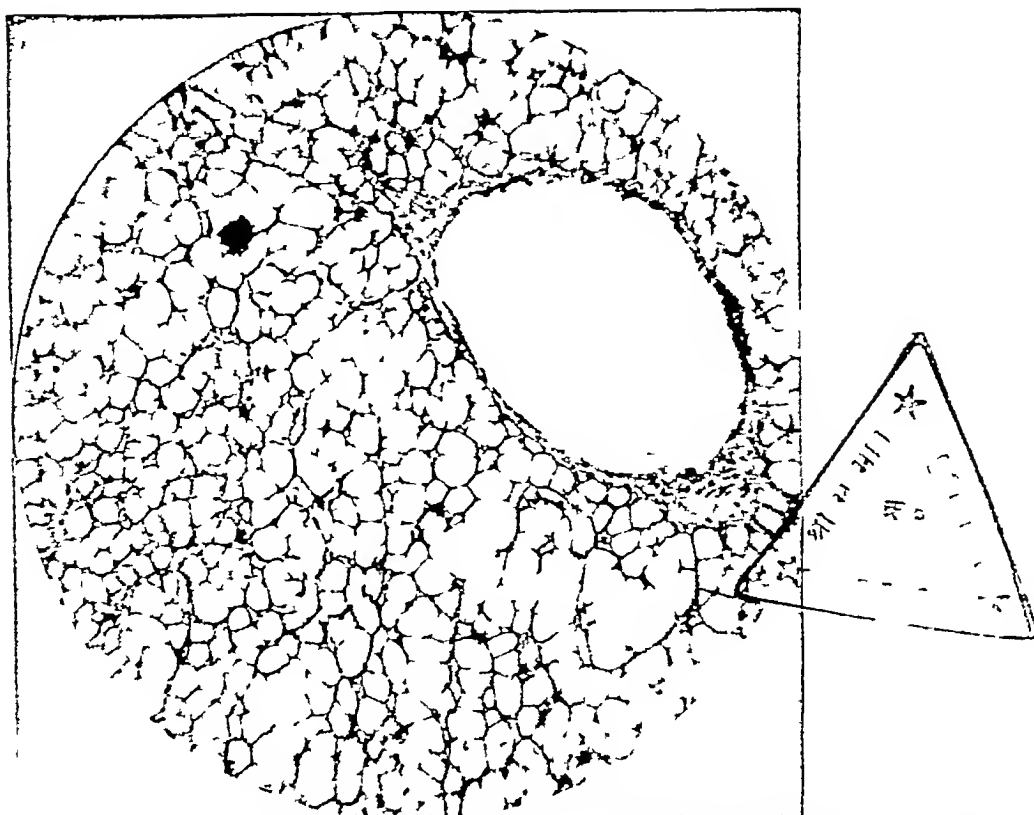


FIG 50 2 Section from liver of depancreatized dog fed a diet low in choline. Ether-soluble material over 60 per cent of wet weight. $\times 200$ (From Best, Huntsman and Young)

the richest known sources of this base. Methionine and vitamin B₁₂ are also present in this extract and if proteolytic enzymes are present they may make methionine available from the dietary protein.

In 1924 Allan, Bowie, Macleod and Robinson in Toronto, and Fisher in Chicago, noted that large fatty livers developed in depancreatized dogs maintained for a long time with insulin. Inclusion of raw beef pancreas in the diet prevented this fatty change. Hershey and Soskin noted a similar protective effect with crude egg lecithin. In 1932 Best, Hershey and Huntsman found that they could produce fatty livers in rats by dietary means in about three weeks as contrasted with a period of more than six months required in the depancreatized dog. With this more rapid method of investigation they quickly identified choline as the active constituent in the lecithin molecule. Betaine was found to exert a similar effect. It was soon discovered that casein and certain other proteins possess lipotropic activity. This activity of protein was largely accounted for in 1937 when Tucker and Eckstein discovered the lipotropic property of methionine. It exerts a lipotropic effect by transferring its methyl group to ethanolamine, thus effecting the biosynthesis of choline (du Vigneaud).

When choline, betaine or methionine labelled in the methyl group with deuterium were fed singly to rats, the choline of tissue phospholipids, the methionine of tissue proteins and urinary creatinine were all found to contain labelled methyl groups. The enzymatic process involved is known as transmethylation and the particular groups concerned are said to be labile methyl groups. Not all methyl groups are labile, a distinction must be made between the processes of demethylation and transmethylation since demethylation does not necessarily give rise to a labile methyl group (e.g. sarcosine).

The original view that the animal body is unable to synthesize methyl groups had to be modified when it was found that rats could survive and grow slowly under certain dietary conditions on rations devoid of any preformed methyl groups (Bennett, Medes and Toennies). Later, the importance of vitamin B₁₂ and of folic acid in the biosynthesis of methyl groups and in the metabolism of the "one carbon fragment" became apparent.

A labile methyl group is not necessary for lipotropic action. Recent studies have indicated that the methyl groups of choline are not labile and *only become so after oxidation* to betaine. It had

been recognized previously that arsenocholine (Welch) and triethylcholine (Channon, McArthur) which contain *no labile methyl groups* exert a lipotropic effect. When fed, they are incorporated into liver phospholipids as the intact molecule. These observations suggest that the lipotropic action of choline is exerted by the intact molecule and not by virtue of its methyl groups. Dietary components with labile methyl groups exert their lipotropic effect by favoring biosynthesis of choline.

The physiological test for a labile methyl group is to feed the compound containing the methyl group to a rat being fed a synthetic diet complete in every sense except that methionine is replaced by homocystine. If growth occurs, which can only happen when methionine is formed, transfer (or synthesis) of methyl groups must have taken place.

When the lipotropic factors² are not available in the diet large amounts of fat accumulate in the liver. In young animals hemorrhagic lesions are seen in the kidneys and other tissues (Griffith and Wade), and these may produce death. The anti-lipotropic factor cholesterol aggravates the liver and kidney changes while choline, betaine and methionine prevent their development. When young rats are subjected for only five days to a diet deficient in choline and its precursors and are then returned to a full normal diet a malignant hypertension of renal origin develops, four or five months later, in a considerable proportion of the animals (Hartroft and Best).

CIRRHOSIS Prolonged exposure to certain toxic agents or dietary deficiencies which cause fatty livers can lead eventually to an excessive production of fibrous connective tissue. The shrunken, distorted liver which results is usually of a peculiar orange to brownish color. It is this aspect which Laennec featured in the name *cirrhosis Kwashiorkor* is a pandemic form of cirrhosis found among

severely undernourished infants and children in many parts of the world. It appears to be related to diets in which the protein is both low in quantity and poor in quality. Some forms of experimental cirrhosis are attributable to choline deficiency. In depancreatized and later in normal dogs Chaikoff produced cirrhosis by prolonged feeding of diets which permitted fat to accumulate in the liver. Later the lesion was studied in rats by György and Goldblatt, Blumberg and McCollum, by Daft, Sebrell and Lillie, by Hartroft and by others. Experimental dietary cirrhosis begins by condensation of fibrous stroma supporting centrilobular cells which have been distended by excessive fat to their bursting points. The ruptured cells atrophy and disappear leaving a scar formed by their stromal remnants. By repetition and extension of the process these scars eventually spread through the rest of the liver including some areas around large portal veins. It can be prevented or cured in animals by the addition of choline or its precursors to the diet. The basal diet should, of course, contain adequate amounts of protein, minerals and vitamins.

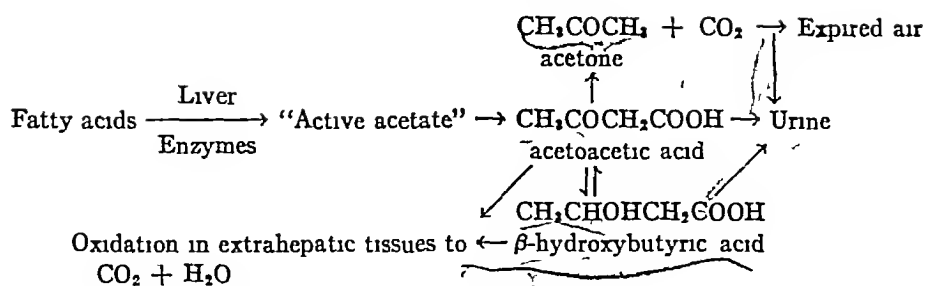
Alcoholic cirrhosis is apparently of dietary, rather than of toxic origin. The excessive consumption of alcohol supplies sufficient calories so that the patient feels no need for other foods which would provide the necessary proteins, vitamins and minerals.

The acute effects of chloroform poisoning on the liver in protein-depleted dogs are diminished or prevented by the administration of methionine (Miller and Whipple). Methionine, cystine or α -tocopherol protect rats against a massive necrosis of the liver which appears when these substances are in large part removed from the diet. These protective effects of methionine are not exerted through its lipotropic action.

In the depancreatized dog choline has been shown to be one of the active components of beef pancreas which as previously stated prevents the development of fatty livers. The protein of pancreas through its methionine content will presumably also exert a lipotropic effect. Furthermore, the pancreatic enzymes will help to liberate choline and methionine from the phospholipids and protein of the diet.

The mechanism by which choline exerts its lipotropic action is unknown but the most hopeful lead is provided by Artom's finding that choline given to animals maintained on a hypolipotropic

² The term lipotropic was originally used to describe the action of choline which prevented the deposit on of, or accelerated the rate of removal of fat from the liver. It now has a somewhat broader meaning and is used to describe the action of choline and its precursors methionine and betaine on fat deposition in liver, kidney and heart. Only those effects of methionine and betaine which are exerted through formation of choline are included in the term "lipotropic". Substances such as cholesterol which cause a deposition of liver fat are referred to as antilipotropic and hypothetical diets free of lipotropic factors may be described as alipotropic. Cystine, under certain dietary conditions increases the deposition of liver fat and the incidence of hemorrhagic kidneys. This is apparently related to an improvement in the diet, the resulting increased food intake stimulates growth and creates an additional need for the lipotropic factors.



Scheme showing the metabolism and routes of excretion of the ketone bodies (James Campbell)

diet increased the oxidation of fatty acids in liver slices from these animals

FAT AND GLYCOGEN IN LIVER While it is true that very small amounts of glycogen are present in very fatty livers and that the fat content is low when large amounts of glycogen are found, it can also be shown that moderate amounts of glycogen can accumulate in moderately fatty livers

INTERMEDIARY METABOLISM OF FAT

During the oxidation of fat to carbon dioxide and water a number of intermediary substances are formed. It has been known for a long time that three substances are produced during the oxidation of fatty acids, namely, acetoacetic acid, β -hydroxybutyric acid and acetone. These are collectively known as the ketone bodies. The discovery of coenzyme A by Lipmann in 1945 and the identification of 'active acetate' as acetyl coenzyme A by Lynen and his colleagues (1951) are achievements of historic importance. We shall attempt to correlate these recent studies with the previously known facts.

Under ordinary conditions the ketone bodies are present in the blood in very small amounts (about 1 to 3 mg of total ketones per 100 ml. blood) and only about 0.3 grams are excreted daily. Ketosis (increased amounts of ketones in blood and urine) may occur when increased amounts of fat are metabolized, for example when the diet is rich in fat, or during fasting when the fat reserves of the body are being rapidly utilized. Certain clinical conditions in which ketosis occurs are diabetes mellitus, fevers, the post-operative state, toxemias of pregnancy, hyperthyroidism, glycogen disease. Ketosis is seen experimentally in pancreatic or phlorhizin diabetes, after administration of certain hormones (growth hormone, adrenocorticotrophic, adrenocortical and, in fasting animals, insulin). The amino acids leucine, isoleucine, phenylalanine, tyrosine and hydroxyproline, produce an increased excretion of ketone bodies in

diabetic animals. Other amino acids may give rise to glucose in the body. It has been estimated that protein predominantly forms glucose rather than ketones. Ketones may be produced through the oxidation of pyruvate by liver slices *in vitro* but in the intact animal the effect of pyruvate appears to be antiketogenic. In general, it may be taken that the precursors of the ketone bodies are the fatty acids.

INTERCONVERSION AND EXCRETION OF THE KETONE BODIES

Acetoacetic acid is the ketone body that is produced first by the tissue cells. Acetone arises by decarboxylation of this acid—a spontaneous reaction that does not necessarily require the enzymatic activity of the tissue cells. The acetoacetic acid may be reduced by the tissues to β -hydroxybutyric acid. The reaction is reversible. In ketosis the greater part of the total ketones occurs as the hydroxy acid. Since β -hydroxybutyric acid is weaker than acetoacetic acid this conversion tends to reduce the unfavorable effects of accumulation of fixed acid.

All three ketone bodies are excreted in the urine. Acetone may also be excreted in the expired air, giving a characteristic odor to the breath of the individual with ketosis. In the days before insulin the diabetic ward could be thus detected at a distance.

SITES OF ORIGIN OF THE KETONE BODIES

It is generally accepted that the liver is by far the most important site of production of ketone bodies. The perfused liver produces abundant quantities of ketones while little comes from other organs such as kidneys, lung and skeletal muscle (Embsen, Snapper, Grünbaum and Neuberger). Himwich and his associates found that liver consistently added ketones, while striated muscle, heart and the organs drained by the portal vein only occasionally added small amounts of ketones,

and sometimes removed them from the blood. It has been observed repeatedly that liver slices may add ketones to the medium and produce relatively large amounts from added fatty acids while other tissues are very much less active. The rate of utilization of acetoacetate by liver tissue is practically nil, i.e. acetoacetate appears to be an end-product of fatty acid oxidation by the liver. In contrast, acetoacetate and β -hydroxybutyrate are readily utilized by muscle and other extrahepatic tissues.

THE MODE OF FORMATION OF KETONES

The higher fatty acids that occur in the body contain an even number of carbon atoms in a straight-chain arrangement. This is not the place to describe the contributions of Knoop, Raper, Schoenheimer and Rittenberg, and many others, all of which pointed to a fundamental two-carbon unit concerned in the metabolism of fatty acids. Two of these units were assumed by MacKay *et al* to condense to form acetoacetic acid. Support for this assumption was provided by Weinhouse, Medes and Floyd by labelling octanoic acid with isotopic carbon in the carboxyl group. The acetoacetate formed when this was incubated with liver slices was found to contain the isotope distributed almost equally between the carbonyl and the carboxyl carbon atoms. This could only have resulted from the condensation of pairs of molecules, each containing two carbon atoms (acetyl

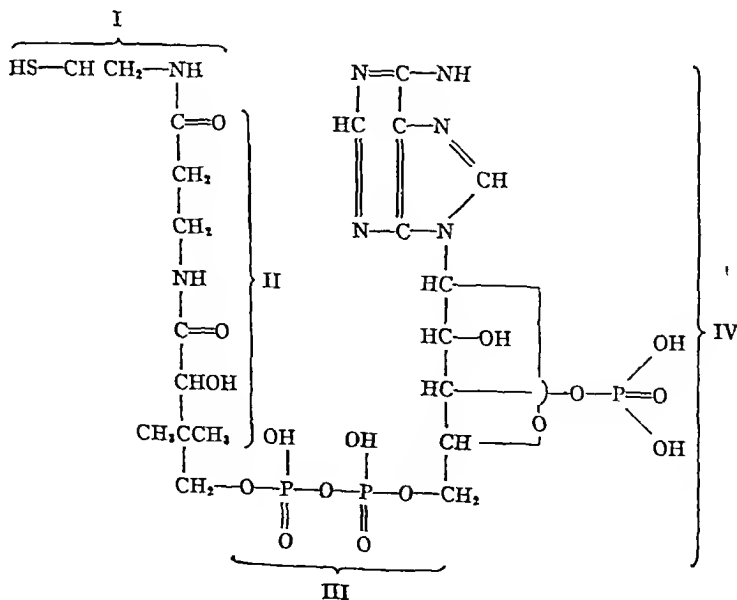
groups) arising by beta-oxidation of the fatty acid. These experiments establish the principle of beta-oxidation and of condensation of "acetate" molecules in the formation of acetoacetate. Subsequent studies have shown that when carboxyl-labelled fatty acids are utilized to form acetoacetate the ratio of the label in the carbonyl carbon label in the carboxyl carbon varies in a regular manner with chain length, rising from about 0.3 for butyrate to 0.8 for octanoate and to 1.0 for higher fatty acids. The reasons for this inequality have been explained by Lynen and by Mahler.

THE ROLE OF THE COENZYME OF ACETYLATION (CoA) IN FATTY ACID METABOLISM

While studying "active acetate", Lipmann investigated the acetylation of aromatic amines such as sulfanilamide by tissues and tissue enzymes and discovered that a heat-stable co factor, the coenzyme of acetylation (CoA) is required.

Coenzyme A has been isolated from liver, which is a rich source, and also from yeast and other microorganisms. It is a dinucleotide consisting of 2-thiol-ethylamine (I) and pantothenic acid (II) linked by a pyrophosphate bridge (III) to adenosine (IV) which has a phosphoryl radical at position 3 of its ribose moiety.

Observations in several laboratories suggested that "active acetate" might be an acetylated derivative of this coenzyme (Lipmann, Stadtman,



Ochoa) The brilliant studies of Lynen and his colleagues led to the isolation from yeast of a substance able to acetylate without the intervention of ATP. This substance proved to be identical with "active acetate." Chemical investigations showed it to be acetyl-coenzyme A, in which the labile acetyl group (i.e. the "active two-carbon unit") is linked to the sulfur atom. To emphasize the presence and importance of sulfur in this high-energy bond the abbreviated formula is often written Ac-SCoA instead of Ac-CoA.

The substrates involved in the enzymic degradation or synthesis of the fats are not the free fatty acids but the activated conjugates with CoA. Coenzyme A acts as an acceptor of the two-carbon units (acetyl groups) formed in the β -oxidation of the fatty acids. The coenzyme combines also with many other organic acids, fatty and otherwise, e.g. propionic, butyric, β -hydroxybutyric, acetoacetic, crotonic, succinic, etc. Thus it may be de-

scribed as an acyl $\left(\text{R}-\overset{\text{O}}{\parallel}{\text{C}} \right)$ acceptor. The sulfur linkage whereby the acyl groups are united

to coenzyme A is a new type of high-energy bond. Acetyl-SCoA may be regarded as analogous in some respects to a mixed acid anhydride and in others to an ester. In spite of the high energy of the thiol-ester bond, these compounds are relatively stable in neutral aqueous solutions at room temperature, a property which fits them for utilization in cellular metabolism. The bond is split and the energy is released in tissues by the activity of specific enzymes. In tissues (e.g. liver, heart, yeast) these fatty acid oxidizing enzymes are bound to the mitochondria of the cells. Ways have been devised to prepare them in soluble form and to separate those involved at each step. The *in vitro* studies were thus made possible.

Energy is required for the first step in the metabolism of the fatty acids, i.e. for the formation of the "activated fatty acid" (CoA Thiol-ester, Reaction A). This energy is provided by a concomitant oxidation of a member of the citric acid cycle. This is the so-called 'sparkling reaction' that initiates the oxidation of the fatty acid. Once the degradation has been 'primed' it proceeds without further addition of energy, two carbon atoms being split off (as acetyl-SCoA) until the acid has been

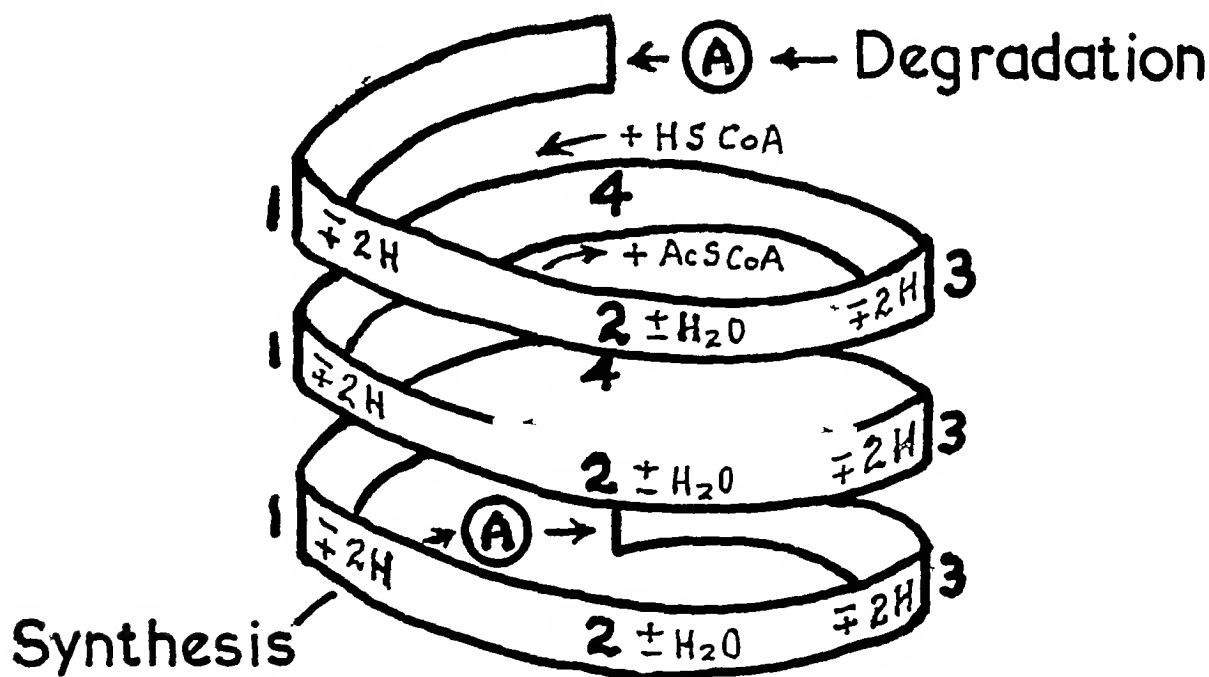
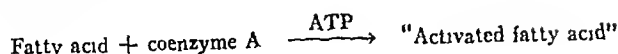


FIG. 50.3 Helical scheme to represent the oxidative degradation and synthesis of fatty acids (Bruno Rosenfeld and C. C. Lucas. With acknowledgement of Lynen's suggestion.)

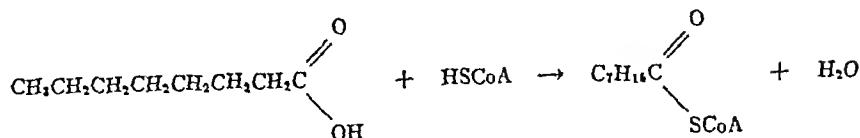
Each complete cycle involves four reactions. In *oxidative degradation* these are (1) dehydrogenation, (2) hydration, (3) a second dehydrogenation, and (4) thiolytic degradation, i.e. elimination of 2 carbon atoms as acetyl coenzyme A following addition of coenzyme A. The *synthetic reactions* occur in the reverse order and in the opposite direction: (1) addition of an active 2-carbon unit (acetyl coenzyme A), (2) hydrogenation, (3) dehydration, and (4) a second hydrogenation. Before the oxidation or synthesis can take place, the fatty acid must be activated (Reaction A) by condensation with coenzyme A. These steps are explained more fully on following pages.

OXIDATIVE DEGRADATION (after Mahler)

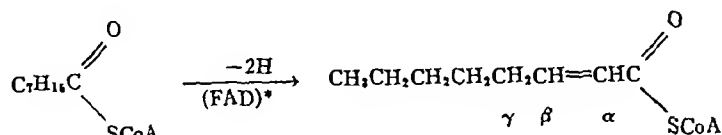
Reaction A Activation Reaction



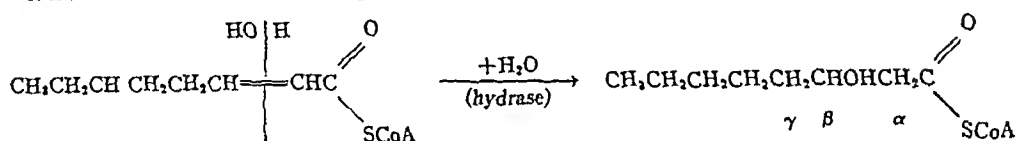
e.g.



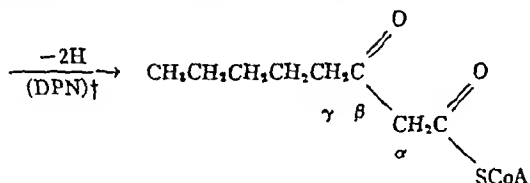
Reaction 1 Dehydrogenation

Enzymatic removal of two hydrogen atoms gives $\alpha\beta$ unsaturated "activated fatty acid"

Reaction 2 Hydration

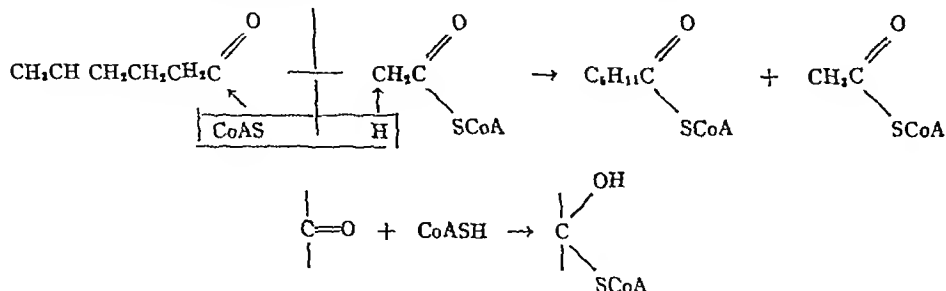
Addition of water at the double bond gives β -hydroxy "activated fatty acid"

Reaction 3 Second Dehydrogenation

Removal of two hydrogen atoms from the β -carbon atom gives the β keto "activated fatty acid"

Reaction 4 Thiolytic Degradation

Coenzyme A adds on at the β -keto position, from this unstable compound an "active two-carbon fragment" (acetyl coenzyme A) is eliminated from the "activated" end of the fatty acid to give a new "activated fatty acid" with 2 less carbon atoms. The molecule of acetyl SCoA thus formed is available to add on at Reaction 4 in the next cycle of the degradation process ‡



then rearrangement occurs with splitting as shown

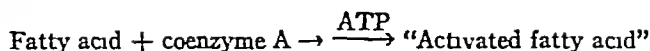
* FAD represents an enzyme derived from flavin adenine-dinucleotide

† DPN represents diphosphopyridine dinucleotide

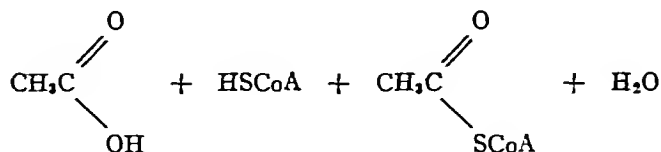
‡ The coenzyme A does not add on as shown in the simplified scheme above. Most likely it first adds on to the carbonyl group

SYNTHESIS

Reaction A Activation Reaction

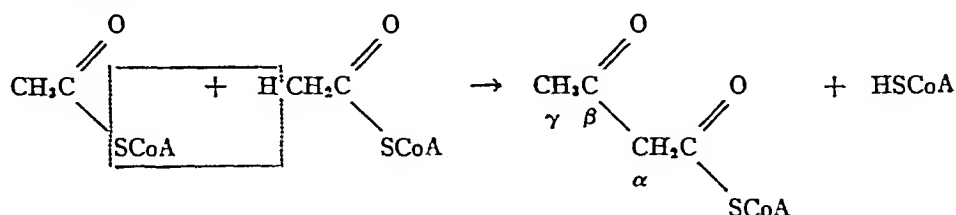


e.g.



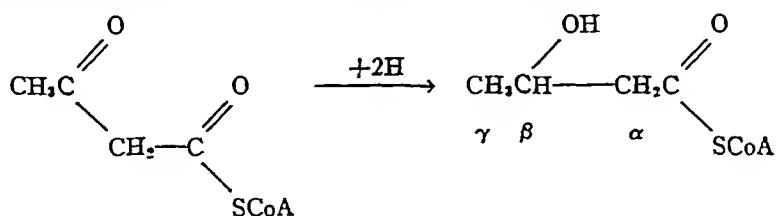
Step 1 (Reverse of Reaction 4) Thio condensation

The "activated fatty acid" condenses with an "active two-carbon fragment" (= "active acetate" = acetyl coenzyme A) to give the activated β -keto acid with 2 more carbon atoms, one molecule of coenzyme A is released to effect this condensation



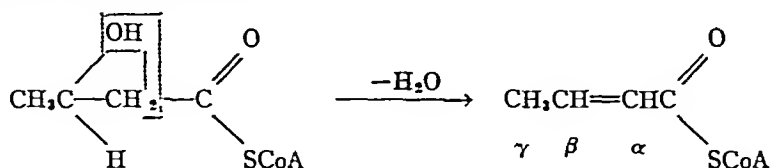
Step 2 (Reverse of Reaction 3) Hydrogenation

Addition of 2 hydrogen atoms to the β -keto activated acid gives the β -hydroxy activated acid



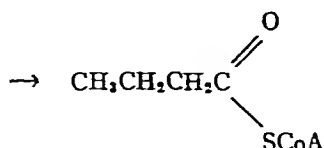
Step 3 (Reverse of Reaction 2) Dehydration

Removal of water gives the $\alpha\beta$ unsaturated activated acid



Step 4 (Reverse of Reaction 1) Second Hydrogenation

Addition of 2 hydrogen atoms to the unsaturated acid finally gives an activated (saturated) fatty acid with 2 more carbon atoms than the acid which entered the cycle.

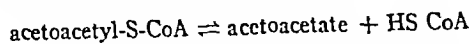


entirely broken down to 2-carbon units. The "active acetate" so formed either enters the citric acid cycle, to be burned to CO_2 and water, or is used in biosynthetic processes. Fatty acid catabolism produces energy, their anabolism requires energy. In both cases the free energy is provided by "active

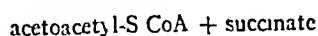
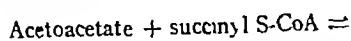
acetate", or more specifically by the bond energy in the acetyl-SCoA linkage.

From a comparison of the turnover number of the individual enzymes, it has been concluded that the β -ketoacyl derivatives, e.g. acetoacetyl-SCoA, may be expected to occur at a much higher

steady state concentration than any other intermediate in the whole enzymatic sequence. Liver, in contrast to kidney and heart, contains a very active deacylase enzyme catalysing the reaction



and a very weak transfer enzyme catalyzing the reaction



Thus we are able to explain why liver tissue gives rise to the formation of acetoacetate as distinct from kidney, heart and muscle tissue (relatively good sources of the transfer enzyme) by which fatty acids are conducted to the catabolic cycle and thus via oxalacetate and citrate to CO_2 and H_2O .

KETOSIS

As mentioned above certain experimental and clinical conditions are characterized by the accumulation of ketone bodies in the blood. Ketosis, in experimental animals, can be attributed largely to increased rate of production rather than diminished utilization of ketones. The ketosis is apparently due to the increased proportion of fatty acids being used by the organism. These are presented to the liver in excessive amounts, causing an accumulation of fat. In both experimental and clinical experience it has been found that when ketosis occurs there is a relative lack of carbohydrate available for use by the liver cells. The administration of carbohydrate produces an antiketogenic effect as do the intermediates of carbohydrate oxidation *in vitro*. Leninger has shown that addition of oxalacetate or succinate to the medium in which liver slices were oxidizing fatty acids decreased the rate of production of acetoacetate and increased that of citrate and other components of the Krebs' cycle. Acetoacetate is an alternative but not an obligatory product of the oxidation of fatty acids by the liver.

There are indications that kidney and other organs may produce acetoacetate, but only in small amounts (Medes, Floyd, and Weinhouse). The enzymatic constitution of kidney and heart provides an explanation for the difference between the ketogenic activity of these tissues and that of liver. The ketone bodies are rapidly utilized by extrahepatic tissues. The rate of utilization is not influenced by the amount of carbohydrate or of insulin present.

Drury and Wick (1953) using C^{14} glucose, have studied the relative rates of oxidation of glucose and β hydroxybutyrate and of glucose and acetate. Insulin increases the oxidation of glucose but not that of the other compounds in eviscerated rabbits. When insulin activity is maximal the administration of hydroxybutyrate or acetate causes a sharp decline in glucose oxidation but has no effect on the rate of its disappearance from the blood. These findings indicate that the cells burn acetate preferentially under these conditions and glucose is displaced from the oxidative sites.

EFFECTS OF KETOSIS

Ketone bodies in small or moderate amounts provide a source of energy for tissues when the supply of carbohydrate is restricted or its utilization is impaired. Acetoacetic and β hydroxybutyric acids formed in liver cells are non volatile and moderately strong acids which must be neutralized by base as soon as they enter the blood stream. In severe ketosis, i.e. when the rate of formation greatly exceeds that of utilization, there may be a dangerous disturbance in acid base equilibrium. These acids are excreted by the kidney and tend to carry base out of the body with them. The kidney attempts to minimize this loss of base by holding it back and thus produces a highly acid urine. A further process to conserve fixed base (K , Na , etc.) is by increasing the formation of ammonia in the kidney. The urine volume is usually high in ketosis. The excretion of these organic acids represents a loss of potential energy. Thus ketosis results in a loss of base, water, nitrogen and energy and leads to acidosis and dehydration.

THE METABOLISM OF ADIPOSE TISSUE

Metabolic changes in the adipose tissue itself have been only infrequently studied. Wertheimer and his colleagues have reported that the adipose tissues of well fed or fasted rats contain little or no glycogen. If, however, after fasting the animals are fed a diet rich in sugars glycogen appears in the adipose tissue. It is not yet established that the glycogen is a step on the road to fat synthesis. The so-called "brown fat" present in the rat and some other mammals is said to be more active in glycogen synthesis than white adipose tissue. Mirsky found that adipose tissue *brevi* contains enzymes which can both synthesize glycogen from glucose-1-phosphate and promote the reverse reaction. The enzyme amylase is also present in adipose tissue. In adipose tissue *in vivo*, insulin promotes the for-

mation of glycogen (Wertheimer and Shapiro, 1948) and of fat (Chernick and Chaikoff, 1951, Wick, Drury et al, 1951)

ABNORMALITIES OF FAT METABOLISM

The problems of starvation and obesity (chapter 51), abnormalities of fat metabolism associated with endocrine disturbances (p 804), and the conditions in which there is an excessive deposition of fat in the liver (p 696) are discussed elsewhere. A profound disturbance of phospholipid metabolism occurs in Niemann-Pick's disease, a condition which is seen in infancy and in early childhood and is invariably fatal. The spleen and liver are enormously enlarged and contain very large amounts of phospholipids. The cholesterol content of these tissues is also increased but not to nearly the same degree. The blood phospholipid is also definitely raised. Histologically there is a great hypertrophy of the reticulo-endothelial system. Very little is known about this disease but it has been suggested that there is some interference with the activity of the enzymes which are responsible for various stages of fat metabolism in the tissues.

In the rare clinical condition involving a disorder of lipid metabolism, Tay-Sachs disease, there is an abnormally high proportion of cerebroside in the brain. Liver, normally, contains little or no cerebroside but in Gaucher's disease the cerebroside content of both liver and spleen is high. The physiological significance of the cerebroside is not known.

Retention hyperlipemia appears to be caused by the sluggish removal of fat from the blood stream. The mechanism underlying the disorder is not clear. Its clinical designation is "idiopathic hyperlipemia." The neutral fat may be increased from 20 to 100 times the normal concentration and total cholesterol may be about double the normal value. It sometimes occurs as an inherited disorder. This retention type of hyperlipemia does not respond to insulin but does to a diet low in fat. The diabetic (transportation) type of lipemia does respond to insulin.

In 1893 Hand described a condition in children which was characterized at autopsy by yellow nodules in the cranial bones and in other situations. Since it will become increasingly difficult to add more names when further cases are described it may be expedient to call the syndrome Hand's disease (Hand-Schuller-Christian, etc.) While these nodules show a predilection for the tissues of the head, i.e., cranial bones, orbit, tubercinereum, they have been noted in most portions of the skeleton (Chester and Kugel). The signs and symptoms

depend on the areas affected. Diabetes insipidus and exophthalmos are common disturbances, i.e., due to lesions near the pituitary gland and in the orbit. Rowland believes that the condition is due to an osseous form of xanthomatosis. These xanthomatous nodules contain large amounts of cholesterol and cholesterol esters, which together may account for a large proportion of their total fat. There is no alteration in the fat content of the unaffected parts of the body. The etiology of the disease is unknown. The generalized xanthomatosis, which is sometimes seen in diabetes mellitus, may disappear when the blood fats are reduced to normal values by appropriate treatment of the diabetic condition.

A great deal of attention is being paid at the present time to the metabolism of fat in tumor tissue but no review will be attempted here. The important part played by cholesterol in the formation of gall-stones has been dealt with (p 546).

Familial hypercholesterolemia is an inherited condition in which there is an abnormally high total serum cholesterol with a normal proportion of free and bound cholesterol. There is frequently an increase in plasma phospholipids. The disease is inherited as an incomplete dominant. The increase in blood cholesterol is not due to dietary intake but to increased production or diminished utilization within the body. There are many reports which indicate that the condition is characterized by a raised incidence of vascular disease.

ATHEROSCLEROSIS

There is now considerable statistical evidence that prolonged hypercholesterolemia and atherosclerosis are associated in some way. The characteristic lesion of an atherosclerotic artery is the cholesterol-rich plaque (containing up to 70 per cent of cholesterol). It is probable that this cholesterol comes from the blood, but this is not proven. The cholesterol in the blood is not present in simple solution, but is carried as lipoprotein complexes, the so-called "giant molecules" (S_r 10-20) of Gofman. While the total serum cholesterol values, or the concentration of the "giant" lipoprotein molecules afford a significant differentiation between *groups* of men who are clinically healthy and *groups* of men who have (or may develop) coronary heart disease, the forecasting efficiency of the values for *individual* diagnosis or prognosis is very low. Ancel Keys and others have shown convincingly that the serum cholesterol level is not readily altered even by large changes in the cholesterol content of the diet. There is, however, a relationship between total serum cholesterol and the total fat content of the diet. Excessive caloric intake rather than cholesterol intake or

ageing *per se* seems to be concerned in the causation of degenerative vascular disease

Diets rich in fat given to dogs in which kidney damage had been produced may result in cardiovascular lesions (McCormick and Holman, 1949) In rats aortic sclerosis may be produced by diets rich in fat when kidneys have been injured by drugs (Lehr and Churg, 1952) Rations extremely low in choline and its precursors may produce severe cardiovascular lesions in young rats fed diets rich

in certain synthetic or natural fats (Stetten, Wilgram and Hartroft) There is a possibility that deficiency of Vitamin E may be important in the production of some of these experimental lesions The bearing of these experimental findings on human atherosclerosis remains unsettled The arguments implicating dietary cholesterol as a causative factor in human atherosclerosis have been well presented by Katz and Stamler (1953) and discussed by Keys (1953) (See p 164)

CHAPTER 51

THE METABOLISM IN STARVATION, SEMISTARVATION AND OBESITY

STARVATION

An animal deprived of food derives energy first from the combustion of its own carbohydrate stores (glycogen). Next, the fat reserves mainly are drawn upon and finally, after these have been exhausted, tissue protein is broken down, the fatty acid part of the molecule is burned while the nitrogen is excreted in the urine mainly as urea. The metabolism of several professional fasters has been investigated. Among the most famous of these are Succi, whose metabolism was studied by Luciani and others, Cette, investigated by Munk and Zuntz, Levanzin by Benedict and Beauté by Cathcart. One of the longest of such fasts upon record is that of Merlatte of Paris which lasted for 50 days. A dog has been starved for 117 days. By the end of this time it had lost 63 per cent of its weight, but was fairly active. Succi on the 40th day of his fast had lost about 25 per cent of his weight. The length of time a man could survive would depend largely upon his physical condition (fat stores, etc.) at the commencement of the fast, but it would probably not exceed 9 or 10 weeks in any event. Terence MacSwiney, Mayor of Cork, after his arrest during the Irish troubles in 1920, went upon a hunger strike which lasted 74 days, it was terminated by his death in coma.

During starvation the loss of weight is not distributed evenly throughout the body, some organs and tissues losing a much greater proportion of their weight than others (fig 51 1). During the first few days the subcutaneous tissues and other fat depots bear the brunt of the effect of the fast. Large quantities of extracellular water are also lost at this time. Later, dissolution of muscular tissue occurs, as indicated by the N/S ratio of the urine (about 14/1). The water lost during this period is derived mainly from intracellular sources. Later, dissolution of protoplasmic structure occurs. The central nervous system, even in prolonged starvation, loses only about 5 per cent of its weight (as estimated from normal standards), whereas the muscles lose about 35 per cent or more. The weight losses of muscular tissue, liver, gastrointestinal tract and spleen run approximately parallel

with that of the body as a whole. The muscle fibers are reduced in size, and many are destroyed. Contrary to general belief, the percentile weight loss of the heart is only a little less than that of the skeletal muscles. The kidney loses only 20 per cent or less of its weight, and the gonads, adrenals and thyroid only from 2 to 8 per cent. Some of the small weight losses reported, e.g., of the central nervous system, may be due to the replacement of solid substance by fluid (see fig 51 1).¹

NITROGEN EXCRETION *The total output of nitrogen in the urine falls for the first day or two of the fast when the body is subsisting chiefly upon its carbohydrate supplies (p. 639). The length of this period varies, of course, with the size of the carbohydrate stores at the commencement of the fast. A steady rise in nitrogen excretion follows, and usually reaches a maximum about the third or fourth day, but from then on it shows a progressive decline and may reach a value of less than 6 grams per day. The nitrogen excreted during the earlier part of the fast is apparently derived largely from the mobilization of "reserve protein" (p. 632). The urea nitrogen excretion at first rises, then falls, its percentage of the total nitrogen excretion also diminishes. The ammonia excretion rises. The creatinine output shows a steady decline but this is largely compensated for by the appearance of creatine (p. 634) so that the creatinine + creatine excretion remains fairly constant. As a result of some experiments of Voit many years ago, it is very often stated that a pronounced rise in nitrogen excretion occurs shortly before death from starvation, which is attributed to an accelerated breakdown of tissue protein. But this so-called *premortal rise* in nitrogen excretion is a very inconstant phenomenon, and of very doubtful significance, for death very often occurs in starved animals in its absence. The total quantity of body protein catab-*

¹There does not appear to be any definite level of emaciation, that is, of weight loss of the body as a whole as a percentage of the normal body weight, at which death is inevitable. Recoveries with judicious feeding and transfusions of plasma have been recorded after losses of body weight up to 50 or even 60 per cent of the normal weight.

olized may be determined by calculation from the total nitrogen excretion on the assumption that tissue protein contains 16 per cent of nitrogen and that practically all the nitrogen derived from the break-down of body protein appears in the urine. That is, each gram of urinary nitrogen represents the deamination of 6.25 grams of protein, so the quantity of protein broken down is calculated by multiplying the figure for the nitrogen excretion by 6.25. On this basis the average daily loss of body protein of an average sized man during starvation is about 50 grams or about 0.4 per cent of the total amount in his body. For a few days following the termination of a prolonged fast the nitrogen excretion shows a pronounced fall—nitrogen is retained for the reconstruction of tissue protein.²

MINERAL METABOLISM. The urinary excretion of *phosphorus* and *sulfur* shows an initial rise, and then a gradual decline, thus showing a curve which roughly parallels that of the total nitrogen excretion. Toward the latter part of the fast the N/P and N/S ratios are around 5.3/1 and 14/1 respectively. The excretion of P in relation to N is greater than one would expect were it all derived from the soft tissues, the same is true for Ca, which indicates that the skeleton contributes to the quantities of these minerals in the urine. The calcium excretion above that which the soft tissues can account for is much greater (10 times) than that of phosphorus, from which it is deduced that calcium carbonate rather than calcium phosphate is liberated from the bones. The ratio of N to S is a little higher than that of the bulk of the soft tissues, which suggests that the sulfur-containing amino-acids are conserved or excreted in only minimal amounts.

The urinary excretion of *chloride*, *sodium*, *potassium* and *magnesium* is reduced from the beginning of the fast. This is to be expected since the intake of minerals is restricted to that provided by the water which is drunk. The concentrations of these minerals in the blood shows little change, but the

sodium bicarbonate is reduced when ketosis supervenes.

KETOSIS. The increase in urinary ammonia (p 461) is a result of the production of excessive amounts of acid metabolites, especially β -hydroxybutyric and acetoacetic acids. The latter are formed as a result of the carbohydrate deprivation, and the consequent excess combustion of fat. Succin toward the end of his fast excreted, daily, from 7 to 13 grams of acetone bodies. In a fasting female subject reported by Folin and Denis the acidosis was extreme, some 18 gram of β -hydroxybutyric acid being excreted upon the fourth day of the fast. Fasting ketosis is much more pronounced in women than in men. This sex difference is not related to the usually greater adipose tissue of females, for a very lean woman excretes larger amounts of ketone bodies than does an obese man. Nor, generally speaking, is ketosis during fasting greater in an overweight person than in one of normal weight. (See also p 704).

CARBOHYDRATE METABOLISM. Even in the later stages of the fast glycogen is found in the liver, and the blood sugar is little depressed. Sugar is apparently synthesized from protein. In the earlier stages there may be a temporary hypoglycemia.

The *basal metabolic rate*, the *body temperature*, *pulse rate* and the *blood pressure* all show a progressive fall throughout the fasting period. See table 66.

SEMISTARVATION, UNDERNUTRITION³

If a diet possessing a caloric value considerably below the energy requirements of the individual is persisted in, as during famine, war blockade, extreme poverty, disease (e.g., stricture of the esophagus or pylorus) or improper feeding of infants, serious nutritional effects result. It must also be remembered that just as an intake of calories over the output will cause obesity so an energy expenditure in excess of the caloric intake will result in a loss of weight. Consequently a man who performs heavy work upon a diet which is adequate only for a sedentary worker will suffer from undernutrition. The economic or other conditions which lead to extreme reductions in the total caloric value of the diet obviously must also cause, as a rule, a reduc-

² The loss of protein varies widely among different organs and tissues. Addis and his colleagues found that in rats fasted for 7 days the several tissues contributed to the total protein loss in the following proportions, muscles and skin 62 per cent, liver 16 per cent, alimentary tract, spleen and pancreas 14 per cent, blood 6 per cent, kidneys 1 per cent, heart 0.5 per cent and the remaining organs 0.5 per cent. Or put in percentages of the protein contents of individual normal organs, the brain loses about 4 per cent of its protein, the muscles, skin and bones about 8 per cent, the heart 18 per cent, the kidneys 20 per cent, the gastrointestinal tract 28 per cent, and the liver 40 per cent.

³ Comprehensive reports of studies on undernutrition have been published within recent years, e.g., *Malnutrition and Starvation in the Western Netherlands*, Netherlands Government 1944-1945, and the two volume monograph by Ancel Keys and his associates, *The Biology of Human Starvation*, Univ. of Minnesota Press, 1950.

tion in the intake of vitamins, essential minerals and first class proteins. As a consequence, the incidence of specific deficiency disorders, e.g., stunting, xerophthalmia, rickets, osteomalacia, scurvy, etc., is also high when the caloric value of the diet is markedly lowered. These special aspects of undernutrition are dealt with in other chapters.

The most pronounced instances of undernutrition and emaciation are met with most commonly in anorexia nervosa, pituitary cachexia, the later stages of malignant disease, and in prolonged and severe infections.

The following is a summary of the main effects of severe undernutrition.

(1) Reduction in body weight—emaciation. The body attempts to make up the caloric deficiency by burning its own tissues. The loss of weight is due chiefly to loss of fat, but also in severe instances to a loss of protein. The nitrogen balance is negative. In children growth is retarded. The positive nitrogen balance is smaller than normal, it may even be negative. In the less severe grades of undernutrition in children the growth impulse continues to cause an increase in height of the skeleton but the muscles and the breadth of the body are poorly developed.

(2) Reduction in basal metabolic rate. The total basal metabolism, i.e., the heat production as related to surface area, is, as one would expect owing to the reduction in the mass of active tissue, diminished in semi-starvation, but the metabolism per unit (kilogram) of body weight is reduced as well. The total metabolism is diminished by nearly 40 per cent, and the heat production per unit of body weight by nearly 20 per cent. The cause of the reduction per kilogram has not been fully explained, but it can be accounted for in part by the decline in cardiac work (by about 50 per cent), the lowered tone of the skeletal muscles, and the subnormal body temperature, which would tend to slow the rate of all chemical reactions in the tissues. The subject is abnormally sensitive to cold, due to the fact that the skin vessels are constricted in an attempt to reduce the dissipation of heat through radiation and convection (p. 731). The skin temperature upon which our thermal sensations depend is therefore lowered. The specific dynamic response to food may be increased.

(3) The subject is readily fatigued, and shows a lack of a zest for physical exertion. Work is performed with the same expenditure of energy as normally, so, though energy is economized in the carrying on of the vital processes as shown by the reduced B.M.R. and cold skin, no economy is effected in the execution of muscular work.

(4) The loss of internal fat which normally serves to support the organs—stomach, kidneys, uterus, etc.—against the effect of gravity, results in their displacement (visceroptosis). The lost body fat is partially replaced by water.

(5) Susceptibility to infections. "Fever and plague

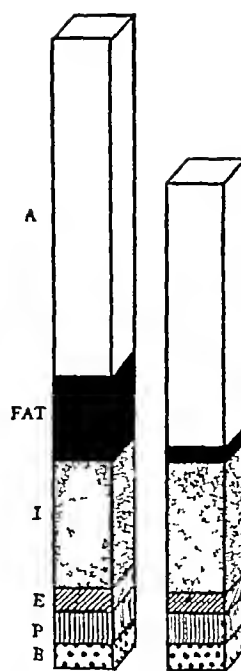


FIG. 51.1 The weights of the major compartments of the body of young men in normal nutrition, on left, after 24 weeks of semi-starvation, on right, A, active tissue (total body weight less the other indicated compartments), B, bone mineral, E, erythrocytes, I, interstitial fluid, (thiocyanate space less plasma volume), from Keys and associates, *The Biology of Human Starvation*, University of Minnesota Press, redrawn and modified.

dog the footsteps of famine." This is an old adage, but with the exception of tuberculosis (in which resistance is definitely lowered) and certain specific effects of vitamin deficiencies there is really little evidence that undernutrition increases the susceptibility to infection.

It is natural to suppose that the production of gamma globulin which is so closely associated with immune bodies would be depressed by undernutrition, especially if involving severe protein deficiency. The investigations of Cannon and his colleagues have gone far to show that this, indeed, does occur. In the Minnesota experiment the concentration of gamma globulin in the plasma of the subjects was moderately reduced, but this was not accompanied by any apparent increase in susceptibility to infections; the semistarved subjects of the experiment had slightly fewer colds over a comparable period than the controls. On the other hand, there is a considerable body of evidence to support the conception that undernutrition actually reduces the susceptibility to certain infections. It has been reported, for example, by McCay that under-fed rats were surprisingly free from the usual laboratory infections; others have concluded from their experiments and observations of naturally occurring disease, that well nourished animals are more susceptible to infection than are undernourished ones. When a severe infection is established, however, the undernourished animal is at a disadvantage, and is more likely to succumb.

It is recognized that some non-infectious conditions

such as hypertension, neoplastic diseases, and diabetes are benefited by undernutrition. In animals, very severe undernutrition definitely inhibits tumor growth, but whether less severe caloric deficiency has such an effect in man is uncertain. The evidence for the incidence of hypertension being reduced is mainly statistical.

(6) Edema not uncommonly (see ch 4)

(7) Psychological changes. Mental apathy, moral deterioration, depression, tendency toward introversion and other changes in personality, and lowered intellectual capacity are common effects. "The brain is in some sort of custody of the stomach and relief of malnutrition gives relief of mental dullness" (Roger Bacon). An unresponsive, complaining and uncooperative attitude in semistarved persons of Western Europe receiving relief after World War II was very frequently observed. Drummond has remarked upon this and states that it disappeared completely when the food intake was raised above 1500-1800 Calories per day.

OBESITY

Some forms of obesity are quite definitely due to a hypothalamic lesion or to a disorder of an endo-

crine organ, pituitary (ch 57), adrenal cortex (ch 59), thyroid or gonads, they will not concern us here. Such origins of obesity are rare and associated with other manifestations of hypothalamic or endocrine dysfunction. Ordinary obesity—the so-called *simple obesity*—is very common and afflicts persons showing no other abnormality.

In many cases of simple obesity there appears to be a genetic element, a conclusion borne out by statistical studies. In surveys of a large number of obese persons, one or both parents were obese, whereas the incidence of obesity in persons both of whose parents are of normal weight is only about 10 per cent. The hereditary obesity of a strain of yellow mice is well known. A Mendelian dominant gene carries the characters for yellowness and obesity, non yellow mice of the same litter are not obese. The basal metabolic rate of the obese animals is said to be reduced and the body temperature subnormal, an hereditary hypothalamic origin of the obesity has been suggested. Another form of

TABLE 69
Subject L. Height, 170.7 cm. Only distilled water was taken during this fast.
(Abridged, after Benedict)

	DAY OF FASTING			
	1st	11th	21st	31st
Body weight, kg	59.60	53.88	50.49	47.39
Rectal temperature at 7 a.m.		36.54	36.04	35.96
Pulse rate, morning, awake	74	61	59	60
Urine				
Total solids, grams	43.51	42.05	31.88	27.07
Total N	7.10	10.25	7.93	6.94
Urea N	5.68	7.66	5.54	4.84
Ammonia N	0.41	1.58	1.57	1.24
Uric acid N	0.112	0.116	0.112	0.122
Creatinine + Creatine N	0.48	0.49	0.38	0.32
Chlorine	3.77	0.36	0.18	0.13
P ₂ O ₅	1.66	1.95	1.60	1.32
N : P ₂ O ₅ ratio	4.28	5.26	4.96	5.26
S	0.46	0.62	0.51	0.49
N : S ratio	15.4	16.5	15.5	14.2
β-oxylbutyric acid		1.4	5.0	4.5
Ca	0.217	0.220	0.237	0.138
Mg	0.046	0.072	0.053	0.052
K	1.630	1.006	0.644	0.606
Na	2.070	0.100	0.066	0.053
Loss of flesh calculated from N loss	213	308	238	208
R : Q, night	0.78	0.72	0.73	0.72
Calories, indirect, twenty-four hours' complete rest	1441	1193	1032	1072*
Calories per square meter (DuBois), twenty-four hours	843	732	653	701†

* Previous day = 1025

† Previous day = 661

obesity in mice which is accompanied by hyperglycemia has been described. The food consumption of these animals is much greater than non-obese controls. An inherent tendency to fatness seems to be illustrated by the common observation that of two persons who appear to eat about the same amount of food, and exercise to the same extent, one may remain thin or of normal weight while the other grows fat. Furthermore, the "spare" person may have a large appetite and remain underweight while the obese may diet himself and still be fat. In order to explain such cases it has been suggested that they are due to an inherited endocrine characteristic. Yet if this were so, some evidence of it should be forthcoming from metabolic studies. On the contrary, the basal metabolic rate per unit of body surface of the subject of the common or simple type of obesity is within normal limits—that is, his energy expenditure at rest is not less than the normal. Nor is work performed more economically than usual by the obese, the reverse is probably true on account of the greater amount of inert adipose tissue.⁴ His greater storage of energy cannot therefore be explained upon this basis. It has been claimed that the specific dynamic action of food (p. 640) is reduced in cases of simple obesity, and it has been suggested that such a reduction, by conserving energy, may play a causative rôle. The smaller specific dynamic action which has been observed could, however, account for no more than a 3 per cent reduction in the total daily metabolism, and is quite inadequate as an explanation. An almost inappreciable increase in the caloric intake or a slightly reduced bodily activity would produce a much greater effect upon the energy exchange. For example, 10 grams of extra fat daily in the diet (e.g., a teaspoonful of butter) yielding 90 Calories, or 23 grams of sugar (about 2 teaspoonsful) would increase the caloric intake of a person of ordinary activity by 3 per cent or so, while a slow walk of a mile would increase the metabolism to a corresponding extent.

Other possibilities have been explored, namely, that the obese person may absorb his food more efficiently, that there exists an inborn peculiarity of the tissue cells, whereby they accumulate fat in excess, or that they release the fat less freely into

⁴Newburgh states that obese persons as compared with the normal produce more heat in the basal state, and expend more energy for the performance of a given amount of work. Their total metabolism measured over 24 hour periods is also considerably greater than the normal.

TABLE 70*

Influence of overweight on mortality in persons aged 45 to 50 years

POUNDS OVERWEIGHT	INCREASE IN DEATH RATE OVER AVERAGE
10	8
20	18
30	28
40	45
50	56
60	67
70	81
90	116

* After Newburgh

the circulation for use as fuel. None of these factors has been found to play a significant part.

It is probable that a hereditary or constitutional factor, in the great majority of instances of ordinary obesity, is more apparent than real, and that a careful investigation of these cases with respect to food intake and muscular activity would reveal a positive energy balance. It is therefore likely that, when obesity shows a familial tendency, the inclination of members of the same family to follow similar habits with respect to diet and exercise, rather than that some inherited endocrine peculiarity, is responsible. Or again, traits which lead to obesity—overindulgence of the appetite, or a distaste for muscular exertion may be inherited. Also, the obese person often, though not a "big eater", indulges in highly concentrated food.

The possibility of a hypothalamic or an endocrine element in the development of obesity is very difficult to prove or disprove. From what is known of the rôle played by the hypothalamus and the endocrines in the control of food intake and in fat metabolism, it is tempting to look in this direction for an explanation of obesity. Obesity might conceivably be due to some hypothalamic or endocrine idiosyncrasy. There is experimental evidence that the hypothalamus exerts an influence upon the hunger sensation (ch. 44), but endocrine effects are more often exerted upon the *distribution* of fat rather than upon the total body weight. There is no *substantial evidence that obesity is due to any other cause than over-indulgence in food in relation to the body's energy requirements*. It must be remembered that the law of the conservation of energy holds for the animal body (ch. 45). Energy taken in the food is either expended or stored as fat and carbohydrate, and to some extent as protein. What-

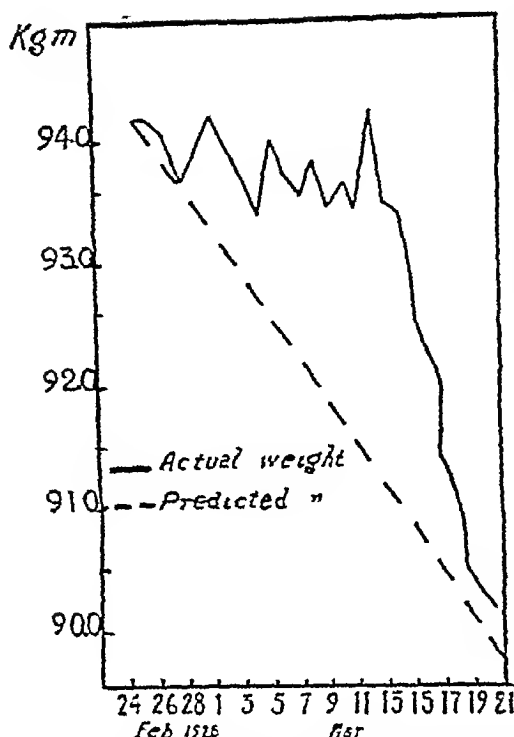


FIG 51.2 Showing the delay in the loss of weight on a reducing diet owing to the retention of the water (After Newburgh modified) This subject lost as much weight in 6 days as it was predicted would be lost in about three weeks as a result of the destruction of body tissue and without water retention.

ever the primary cause may be, whether hypothalamic or endocrine disease, or whether it is due simply to overeating obesity is the result of an energy intake in excess of the energy output. Theoretically, an energy balance could be struck by increasing the exercise of an obese subject. But increasing the energy output by exercise is much less effective than dietary restriction, physical exertion increases the appetite and the subject finds it difficult to resist satisfying it.

THE "COST" OF OBESITY (See table 70) (1) Owing to the increased weight, muscular exertion places a greater load upon the heart and circulatory system. The incidence of arterial hypertension is relatively high in persons who are overweight. (2) Dissipation of heat by conduction and radiation (ch 53) is reduced through the heat-insulating effect of the mantle of subcutaneous fat. Sweating is in consequence more profuse. (3) Diabetes is more common in the obese than in persons of normal weight. Joslin refers to diabetes as "the fat man's folly", and to obesity as "the open door to dia-

betes" Reducing the weight of an obese diabetic has been found to diminish the sugar excretion by as much as 75 per cent. (4) Persons who are grossly obese are said to be less resistant to infections and poorer surgical risks than those of normal weight. (5) The incidence of gallstones is relatively high in the obese according to Baumann, 88 per cent of persons with gallstones are overweight. (6) *Life expectancy* Insurance statistics show that overweight after the age of 35 years is associated with a death rate much higher than that of lean persons or of those of normal weight, a fact which has been pithily expressed in the phrase, "the longer the belt the shorter the life."

GENERAL PRINCIPLES IN THE TREATMENT OF OBESITY The basis of treatment should be, as already indicated, (a) reduction in the caloric intake and (b) increase in the energy expenditure through moderate exercise. The dietary restrictions should not be extreme and should be made gradually. The extent to which the caloric intake should be reduced will depend upon the degree of obesity and the amount of exercise prescribed, but it should never be pushed to the point where the protein of the subject's tissues is drawn upon for energy purposes, nitrogen equilibrium should be maintained. Above all things a properly balanced diet should be devised and an adequate supply of vitamins and minerals provided. When the obesity is pronounced (body weight 25 per cent or more above normal) weight reduction is brought about by placing the subject upon a diet possessing a caloric value 40 or even 60 per cent below his energy requirement, he is thus forced to consume his own fat. The caloric value of human adipose tissue is about 3500 Calories per pound.⁴ The total requirement of the average obese subject is around 2500 Calories per day. A reduction in the energy intake by 40 per cent of the requirement, that is, to 1500 Calories, will therefore entail a weight loss of nearly a third of a pound per day. The subject is kept upon the restricted diet until the desired weight has been reached, his caloric intake is then adjusted to his requirement (see ch. 56).

The failure of an overweight person to lose weight after having been placed upon a "reducing" diet has been a mystifying observation which appeared to refute the conclusion that obesity is due simply to a balance of energy intake over energy output. But the original weight is maintained for only two or three weeks, and is due to the retention of water (p 23) which replaces the tissue broken down for energy purposes. At the end of this period water excretion increases above the normal and the weight falls rapidly (Fig 51.2)

⁴ Pure human fat has a caloric value of approximately 9.5 Cal per gram. 3500 Cal per pound here given is the value after allowance has been made for connective tissue and water content.

In planning a weight-reducing diet, it is toward the restriction of the more concentrated forms of food that attention is particularly directed. It will be recalled that the caloric value of fat is more than twice that of carbohydrate. Also fatty foods, since they contain little or no water, are more concentrated than starchy materials. Lard, dripping, olive oil, etc., are 100 per cent fat, whereas the protein and carbohydrate of white bread amount together to only about 60 per cent of its fresh weight. Ordinary fats and oils should therefore be restricted. Butter, though about 85 per cent fat, should not be disallowed owing to its value as a source of vitamin A. Sugar in the form of sweetening for beverages, in jams, honey or chocolates is a highly concentrated food. By curtailing its consumption a large reduction in the caloric value of the diet can be effected without the disadvantages attending the reduction of some of the

other foods. Bulky foods, e.g., green vegetables and salads of low caloric value but satisfying to the appetite may be substituted. Undue restriction of the water intake is sometimes practised but this measure appears to be of no benefit and may be a detriment to health. Alcohol has a high caloric value and is therefore, except in minimal amounts, excluded from the diet. *Thyroid extract* is sometimes employed to raise the metabolic rate and so reduce the obesity. The hormone is clearly indicated as a means of raising a low metabolic rate to normal, otherwise its use is not to be advocated except in exceptional cases and only when the subject's basal metabolic rate can be followed by frequent determinations. *Dinitrophenol* is another agent which raises the metabolic rate and has been employed in the treatment of obesity. It is a dangerous drug unless given under the strictest supervision. Cataract has been reported following its use, liver injury may also result.

METABOLISM IN MUSCULAR ACTIVITY

The Structure of Skeletal Muscle

A typical skeletal muscle such as the gastrocnemius or biceps is composed of large numbers of elongated cells, usually referred to as muscle fibers. The fibers are bound together into bundles or fasciculi of different sizes which are grouped again into larger masses to form the whole muscle. A connective tissue sheath of reticulated fibers, called the *perimysium*, encloses each fasciculus and separates it from its neighbors. The connective tissue sheath enveloping the whole muscle is called the *epimysium*. The muscle fibers of different muscles range in length from 1 mm, in a minute muscle such as the stapedius, to about 4 cm in a long muscle such as the sartorius. The cells traverse the entire length of the shorter fasciculi but in the longer muscles they are joined together by tapering ends the tips of which are applied side to side, to form elongated chains. This is the commonest arrangement, but in some muscles designed for maximum contractile force with a minimum of shortening the fibers are disposed in parallel instead of in series or they may run diagonally to a central tendon. The muscle fibers also vary widely in diameter, which measures about 10 microns in the small muscles to 100 microns in the larger.

The body of the cell contains a limited and variable amount of material which corresponds to the cytoplasm of other cells, and is called the *sarcoplasm*. The latter forms a weak gel which surrounds large numbers of fine threads packed closely together and running from end to end of the cell. These are the *myofibrils*, they have a diameter of from 1-2 microns. Under the high power of the microscope they appear as a faint pattern of parallel lines. At the insertion of the muscle they become continuous with the fibrils in the tendon. This longitudinal striation is also seen in smooth muscle. But the most striking and characteristic microscopical appearance of skeletal muscle fibers is their *cross striation* or *stripes*. These are alternating light and dark bands which run across the width of the fiber. When the myofibrils are separated by microdissection, the cross-striations are seen to be a feature of the fibrils themselves, the cross stripes of the whole fiber are the result of the accurate alignment of the dark and light bands from fibril to fibril across the breadth of the cell. The dark and light sections of the fibers are known as the Q (or A) and I (or J) bands, respectively. The Q band in the relaxed fibers is from 3 to 6 micra long, and is a little broader than the I band. It stains deeply with the usual stains while the I band remains pale, but even in unstained sections of muscle the Q band is dark as compared with the adja-

cent I band. When the fresh muscle fiber is viewed by polarized light, the Q band appears bright and the I band dark. This is due to the former being doubly refractive (DR). The I band has a single refractive index. On account of its birefringent property the Q band is also known as the *anisotropic* band or disc, while the I band is termed *isotropic*. Both the Q and the I bands are bisected by narrow cross lines. The line in the Q band is bright and known as the H line (or Henson's line), that in the I band is called the Z line, Dobie's line or Krause's membrane. The Z line is not confined to the myofibrils but in teased specimens can be seen to pass through the sarcoplasm between them, and appears to be of a membranous nature. It extends to the sarcolemma to which it becomes attached. When the muscle contracts the Q band or disc shortens, whereas the I band shows little or no change.

Each muscle cell contains several hundred nuclei which in vertebrate skeletal muscle are situated, for the most part, superficially just beneath the *sarcolemma*. The latter is a delicate, transparent, but tough membrane which envelops the entire cell. It is composed of interlacing collagenous fibrils embedded in a clear matrix of colloidal material. The fibrils are continuous with those in the endomysium.

When as a result of growth or athletic training the muscles enlarge and become more powerful, they do so not by increasing the number of their constituent fibers but by an increase in size of the individual fibers. The fibers are also larger in persons who are well nourished as compared with those in states of malnutrition. The difference in size of the fibers in the two instances is due to the amounts of sarcoplasm which they contain.

The foregoing is a description of the muscle fibers as seen under as high a magnification as is possible with the ordinary microscope. But, when observed with the electronmicroscope, it is found that the myofibrils are themselves made up of still finer longitudinal filaments. These filaments or *micellae*, which are the ultimate contractile structures, are composed of two proteins (p. 715) and have a diameter ranging from 25 to 50 Ångström units (Å).

According to Szent György, the distinctive optical properties of the two segments (Q and I bands) of the fibrillae are not due, as has been generally supposed, to differences in their composition or structure. Gerandus and Matoltzky found that when muscle was treated with reagents which dissolved the myofibrils the muscle retained its cross striation but a curious change—a reversal of the optical properties of the bands—oc-

curved. The Q bands were now isotropic and the I bands doubly refractive (anisotropic), but unlike the normal *positive* birefringence of the Q bands, the double refraction (DR) of the I bands was negative. From this demonstration it is believed that the normal cross-striation of muscle is due to a protein material which surrounds the micellae and has a discontinuous or periodic distribution. This substance has a negative DR which in the intact muscle fiber just balances at regular intervals the uniform positive DR of the micellae, thus causing the isotropism exhibited by the I segments. In the intervals between the I segments the filaments or micellae exhibit unopposed their positive DR property, and thus give rise to the Q segment. On account of its negative DR the interfilamentous protein has been named N-protein. It is a nucleoprotein. An advantage is seen in the sectional distribution of the N-protein in that when the filaments shorten and thicken, spaces are available into which the material surrounding them can be displaced. Otherwise the contraction would be resisted and impeded by the presence of this material.

The Chemical Constitution of Muscle

Muscle is composed mainly of protein (18–20 per cent) and water (75–80 per cent). The remainder of its substance is made up of various minerals and a large number of organic compounds some of which are found exclusively in muscular tissue. The chief electrolyte is potassium which is present in a concentration of about 400 mg per 100 grams. *Magnesium, calcium* and *sodium* are found in much smaller amounts. Muscle contains a great variety of substances generally grouped under the heading of nitrogenous extractives, such as *creatine, phosphocreatine, alanylhistidine, adenosinetriphosphate, creatinine, purine bases, carnosine, anserine* (bird and fish muscle) etc., etc.

Muscle contains between 0.5 and 1.5 per cent of glycogen, from 0.02 to 0.04 per cent of glucose and about 0.01 per cent of lactic acid. Phospholipids (1.0 per cent), cholesterol (0.07–0.18 per cent), vitamins, enzymes and a host of other organic materials are present, many of which will be mentioned later.

Of the proteins of muscle, *myosin* and *actin* compose the contractile elements, the filaments or micellae. Myosin makes up about 60 per cent of the total protein and actin about 12 per cent. The role which they play in muscular contraction will be dealt with later. The other proteins which enter into the construction of the sarcoplasm are N-protein, a nucleoprotein already mentioned, *globulin X, myogen*, and *myoalbumin*. The oxygen-holding pigment *myoglobin* (p. 58) in red-muscle is about 1 per cent.

THE CHEMICAL PHYSIOLOGY OF MUSCULAR CONTRACTION

The mechanical response of an isolated muscle to stimulation is not *accompanied* by an increased consumption of oxygen, extra oxygen is not consumed until after contraction and relaxation are over. So, there are two phases in the contraction cycle, an *anaerobic (anoxidative) phase* and an *aerobic (oxidative) or recovery phase*, during which the muscle is restored to its previous state. If the muscle be stimulated repeatedly in an atmosphere of nitrogen it contracts forcibly at first, but soon becomes fatigued, since it cannot obtain the oxygen necessary for its recuperation between the individual contractions. Lactic acid accumulates and the muscle enters into rigor. If, at the onset of fatigue, oxygen is re-admitted, the lactic acid disappears, and the muscle recovers its original power to contract. The lactic acid concentration at which complete fatigue of skeletal muscle ensues (lactic acid maximum) is from 0.3 to 0.6 per cent. The lactic acid is derived from the breakdown of glycogen (ch. 49). As the lactic acid concentration rises the carbohydrate stores diminish. Yet, the onset of fatigue is not due to the exhaustion of the glycogen stores, for the muscle fails to contract before the latter have disappeared. It is more likely that the high acidity inhibits the enzymes through whose action glycogen breakdown is brought about. Phosphoric acid also accumulates in a muscle contracting in the absence of oxygen. The phosphoric acid production rises rapidly during the earlier contractions, soon reaches a maximum and then ceases.

That a muscle can contract anaerobically has been known for many years. It has also long been known that lactic acid and CO_2 are produced by a muscle contracting in nitrogen. Spallanzani 150 years ago observed that snails placed in nitrogen evolved CO_2 . In spite of these earlier observations it was thought, nevertheless, that a muscle derived its energy from oxidative processes—for how otherwise was the CO_2 produced? In order to explain the phenomenon it was supposed (Hermann, Pflüger) that an oxygen store was contained in some giant molecule (termed “inogen” by Hermann) within the muscle substance itself. So “intramolecular” oxygen was spoken of as the hidden oxygen reserve from which the muscle drew for its anaerobic contraction. Lactic acid also, it was thought, was derived from the breakdown of this hypothetical molecule. This theory was disproved by the classical experiments of Fletcher and Hopkins in 1907. These observers showed that the CO_2 which appeared during

the anaerobic contraction was not the result of oxidation but was *preexisting* CO_2 —i.e., simply CO_2 which had been liberated by the action of lactic acid upon sodium bicarbonate in the muscle fluids. They showed that the oxidative processes occurred after the contraction was over, that lactic acid then disappeared and CO_2 was formed. It was erroneously supposed at this time, that the lactic acid which disappeared had been *completely* oxidized to CO_2 and water. That glycogen was the lactic acid precursor was indicated by the fact that the appearance of lactic acid in muscle fatigued in nitrogen was proportional to the glycogen loss. Also, an RQ of around unity, found later by Meyerhoff, for the recovery phase of isolated muscle, indicated that its fuel was carbohydrate.

Meyerhof also showed that when oxygen was admitted to a fatigued frog's muscle, the amount of gas consumed was only one-fifth of that expected if the disappearance of the lactic acid were due to its oxidation to CO_2 and water. The heat produced during the oxidative phase was also much less than it should be, were all the lactic acid burned. It was found, further, that glycogen *increased* in the fatigued muscle recovering in oxygen.

Emhden had previously shown the importance of phosphate in the activity of muscle. Increased excretion of phosphate in the urine occurs as a result of muscular exercise, and Emhden found that when muscle juice was incubated with a solution of bicarbonate, lactic acid and free phosphoric acid appeared in nearly equimolecular amounts. He suggested that the *immediate* precursor of lactic acid was a hexose phosphate. This he termed *lactacidogen*. A hexose phosphate had been shown by Harden and Young to be formed as an intermediary in the fermentation of sugar by yeast. When Emhden added this ester to muscle juice an increased formation of lactic and phosphoric acids occurred. He also claimed to have demonstrated an increased production of phosphoric acid in a muscle during its contraction. The phosphoric acid as well as the lactic acid was believed to be derived from the breakdown of *lactacidogen* (i.e., hexose phosphate).

THE CHEMICAL CHANGES OCCURRING IN MUSCLE CONTRACTING IN THE ABSENCE OF OXYGEN, PHOSPHORYLATION, GLYCOGENOLYSIS AND GLYCOLYSIS

(See also chapters 32 and 49)

Muscle does not depend upon immediate oxidative processes to liberate the energy necessary for its contraction. It will continue to contract for some time in the complete absence of oxygen. The energy stored in glycogen is thus liberated anaerobically. But for the recovery of the muscle, that is, for the reaccumulation of its energy stores,

oxygen is required. The chemical changes involved in the entire contractile process are thus divided into two phases, the *anaerobic* or *contraction phase* and the *aerobic, oxidative* or *recovery phase*. The muscle is comparable to a machine, such as a submarine engine which runs when submerged upon stored energy supplied in electric batteries. The batteries are recharged when the submarine comes to the surface. Expressed in another way, the muscle when it contracts runs up an oxygen debt (p. 725) which it pays when the contraction is over.

Evidence for the anaerobic chemical changes about to be outlined has been obtained very largely from studies of the enzyme systems in aqueous extracts of muscle and in yeast juice during alcoholic fermentation.

One of the first steps in the liberation of energy for contraction is the breakdown of glycogen stored in the muscle fiber. The glycogen takes up inorganic phosphate (H_2PO_4) and splits off simultaneously *glucose-1 phosphate* (Cori ester) molecules. This reaction is catalysed by *phosphorylase* (see table 71). The cleavage of the glycogen molecule occurs successively at the 1-4 C—O—C linkages between the glucose units. In the separation of the latter, an H atom of the phosphate group becomes attached to the 4 carbon atom, while the 1 carbon atom of the adjacent glucose unit holds the —O— PO_3H_2 group. The uptake of phosphate and its cleavage into glucose phosphate units has been aptly called *phosphorolysis* by Parnas, since, obviously, it is analogous to hydrolysis in which water (H—OH) instead of phosphate is taken up, the H atom and OH atoms becoming attached, respectively, to the adjacent groups of the larger molecule (protein, disaccharides, poly-saccharides, etc.) when they separate (see table 71).

The glucose-1-phosphate formed in the phosphorylatic reaction undergoes an intramolecular transference of its phosphate group to the 6 carbon atom, being thus converted to *glucose-6-phosphate* (Robison ester). This reaction is catalyzed by the enzyme *phosphoglucomutase*. The glucose-6-phosphate, through the action of *phosphohexose isomerase*, gives rise to fructose-6-phosphate (Neuberg ester), which then receives a phosphate group from adenosinetriphosphate (ATP), to form fructose-1-6-diphosphate and adenosinediphosphate (ADP).

The fructose-1-6-diphosphate is acted upon by *aldolase* and split into the triosephosphates,

dihydroxyacetone phosphate and *3-phosphoglyceraldehyde* (Fischer-Baer ester) The 3-phosphoglyceraldehyde reacts with inorganic phosphorus, and in the presence of *triosephosphate dehydrogenase* and DPN (diphosphopyridine nucleotide or Coenzyme 1) is converted to *1,3-diphosphoglyceric acid* DPN accepts H_2 and is reduced to $DPNH_2$. As rapidly as dihydroxyacetone phosphate is produced it is transformed to 3-glyceraldehyde phosphate (called the "reactive form" of the sugar) which in the presence of *triosephosphate dehydrogenase* is rapidly removed. The reaction is thus "drawn to the right"

The 1,3-diphosphoglyceric acid undergoes dephosphorylation to form *3-phosphoglyceric acid*, the released phosphate group converting ADP to ATP. By intramolecular transfer of the phosphate group 3-phosphoglyceric acid is converted to *2-phosphoglyceric acid*, the reaction being catalyzed by *phosphoglyceric mutase*. 2-Phosphoglyceric acid is converted by dehydration to *phospho-enolpyruvic acid* through the action of *enolase*.

The phospho-enolpyruvic acid reacts with ADP to form *pyruvic acid* and ATP. This reaction is catalyzed by a *transphosphorylase* in the presence of magnesium and potassium ions.

Pyruvic acid in the absence of oxygen is reduced to *lactic acid* by the reduced coenzyme ($DPNH_2$) in the presence of *lactic dehydrogenase*, the coenzyme being restored to the oxidized form (DPN).

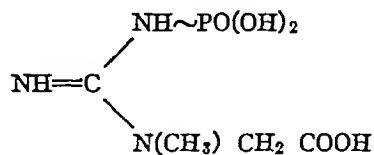
Under aerobic conditions the reduced coenzyme is not oxidized by pyruvic acid, as shown in the last reaction, but by molecular oxygen through the combined action of a specific flavoprotein and the cytochrome system (ch 32). The result is that with an adequate supply of oxygen no lactic acid is formed, and of that produced anaerobically $\frac{1}{2}$ is oxidized through pyruvic to carbon dioxide and water, and the remainder resynthesized to glycogen. In the intact animal synthesis occurs in the liver.¹ The oxidation of lactic acid to pyruvic acid is brought about through the action of *lactic dehydrogenase* in the presence of DPN. In the presence of oxygen the pyruvic acid is oxidized in a series of steps called the *tricarboxylic acid cycle*. Though there is not unanimity as to the details of the chemical changes involved in the dissimilation of pyruvic acid, the researches

of a number of workers, especially Szent Györgyi, Krebs, Werkman and Wood and Evans and Slotin, permit the scheme to be drawn as in table 72, p 721. Thus oxaloacetic acid and acetyl-coenzyme A undergo condensation to form citric acid. The citric acid loses a molecule of water and is converted to *cisaconitic acid*, this then regains a molecule of water to produce *isocitric acid* which is oxidized to *oxalosuccinic acid*. The latter is decarboxylated to yield α -ketoglutaric acid which is oxidized to succinic acid, carbon dioxide and water. The succinic acid is oxidized through fumaric and malic acids to oxaloacetic acid. With the regeneration of oxaloacetic acid the cycle repeats. At each turn of the cycle an acetyl group derived from pyruvate is oxidized, each molecule of pyruvic acid, therefore, yields two molecules of water and three molecules of carbon dioxide, while a molecule of oxaloacetic is reformed. A single molecule of oxaloacetic acid can be used in this way repeatedly in the oxidation of acetyl groups to carbon dioxide and water, it, therefore, acts in the manner of a catalyst.

The energy which the anaerobic production of lactic acid from glycogen yields for the performance of work is only a small fraction of that which results from the oxidation of lactic acid. For each gram mol. of lactic acid produced the free energy amounts to about 29,000 calories, whereas some 325,000 calories are liberated by the oxidation of a corresponding amount of lactic acid to carbon dioxide and water.

THE ROLE PLAYED BY PHOSPHAGEN (PHOSPHO-CREATINE OR CREATINE PHOSPHATE)

In 1927 Eggleton and Eggleton obtained a creatine-phosphorus compound from muscle which they named *phosphagen*. About the same time Fisk and Subbarow isolated the same substance and called it *phosphocreatine*, it is also known as *creatine phosphate*, or briefly CP. It has the following formula



Phosphoarginine is the corresponding phosphagen in invertebrate muscle. In the muscles of some species both compounds are present.

¹ According to Meyerhof resynthesis of glycogen from lactic acid occurs in the isolated frog muscle, but this probably does not occur in intact mammalian muscle.

Contraction without lactic acid production

An experiment of Lundsgaard (1929) revealed the essential importance of phosphocreatine in the contractile process. He found that after the injection of sodium iodoacetate into the dorsal lymph sac of a frog, violent contractions of the muscles occurred followed by rigor. When a muscle of the poisoned animal was isolated and stimulated electrically in nitrogen it responded, but *no lactic acid was produced*. After about 100 contractions, it became fatigued. This is much sooner than the onset of fatigue in normal muscle, and, unlike the latter, the poisoned muscle became slightly *alkaline* in reaction. About 0.5 mg of lactic acid per gram of tissue is formed in a normal muscle after a similar number of contractions.

Upon analysis the poisoned muscle shows a loss of glycogen, and of phosphocreatine, and an increased content of hexose phosphate. The lactic acid and phosphocreatine phosphorus values of normal and poisoned muscles after 150 contractions each are shown in the following table (after Lundsgaard)

		Lactic acid mg per cent	Phosphagen phosphorus mg per cent
Normal	Resting	25	61
	Working	84	40
Poisoned	Resting	16	57
	Working	15	0

It was concluded that the disappearance of CP was not due to its being used up more rapidly, but to its failure to be resynthesized. The energy for resynthesis was presumably derived from glycogen—lactic acid breakdown (glycolysis). The poison destroys the action of triosephosphate dehydrogenase, glycolysis being, therefore, arrested at the stage where 3-phosphoglyceric aldehyde is converted to 1,3 phosphoglyceric acid. Thus, iodoacetate inactivates the carbohydrate mechanism which is essential for driving the phosphagen cycle.

Creatine phosphate does not furnish directly the energy for contraction, as was once supposed, but through its reaction with ADP is a source of energy for the resynthesis of ATP. Thus, $CP + ADP \rightarrow C + ATP$. This is known after its discoverer as the Lohmann reaction. Thus, the energy-rich phosphate bond of CP, as shown by the symbol \sim in the formula, and which amounts to about 10,000 calories per gram mol, is transferred to ATP. The latter is believed to be the immediate source of the energy for contraction.

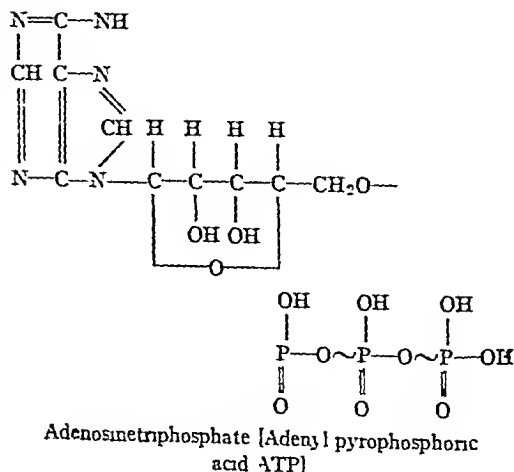
The energy for the resynthesis of phosphagen is furnished by the energy-rich phosphate bonds generated in glycolysis, the transfer is effected through ATP, by a reversal of the Lohmann reaction, i.e., ATP yields phosphorus to creatine, creatine phosphate and ADP being formed. Phosphocreatine resynthesis occurs more readily in the presence of oxygen than under anaerobic conditions, the muscle, therefore, becomes fatigued much sooner in the absence than in the presence of oxygen.

When the muscle has been poisoned by iodoacetate resynthesis of ATP through the Lohmann reaction proceeds for a time, but, since glycolysis has been arrested, CP ceases to be resynthesized, and after a series of contractions, the supply is exhausted.

From the foregoing observations it may be concluded that the function of phosphocreatine is to hold a reserve of phosphate-bond energy which is readily available for transfer to the ADP/ATP system in the earlier stages of contraction. Later, energy for contraction is transferred to ATP from phosphate-bonds generated in glycolysis.

THE ROLE PLAYED BY THE ADENYLIC ACID SYSTEM

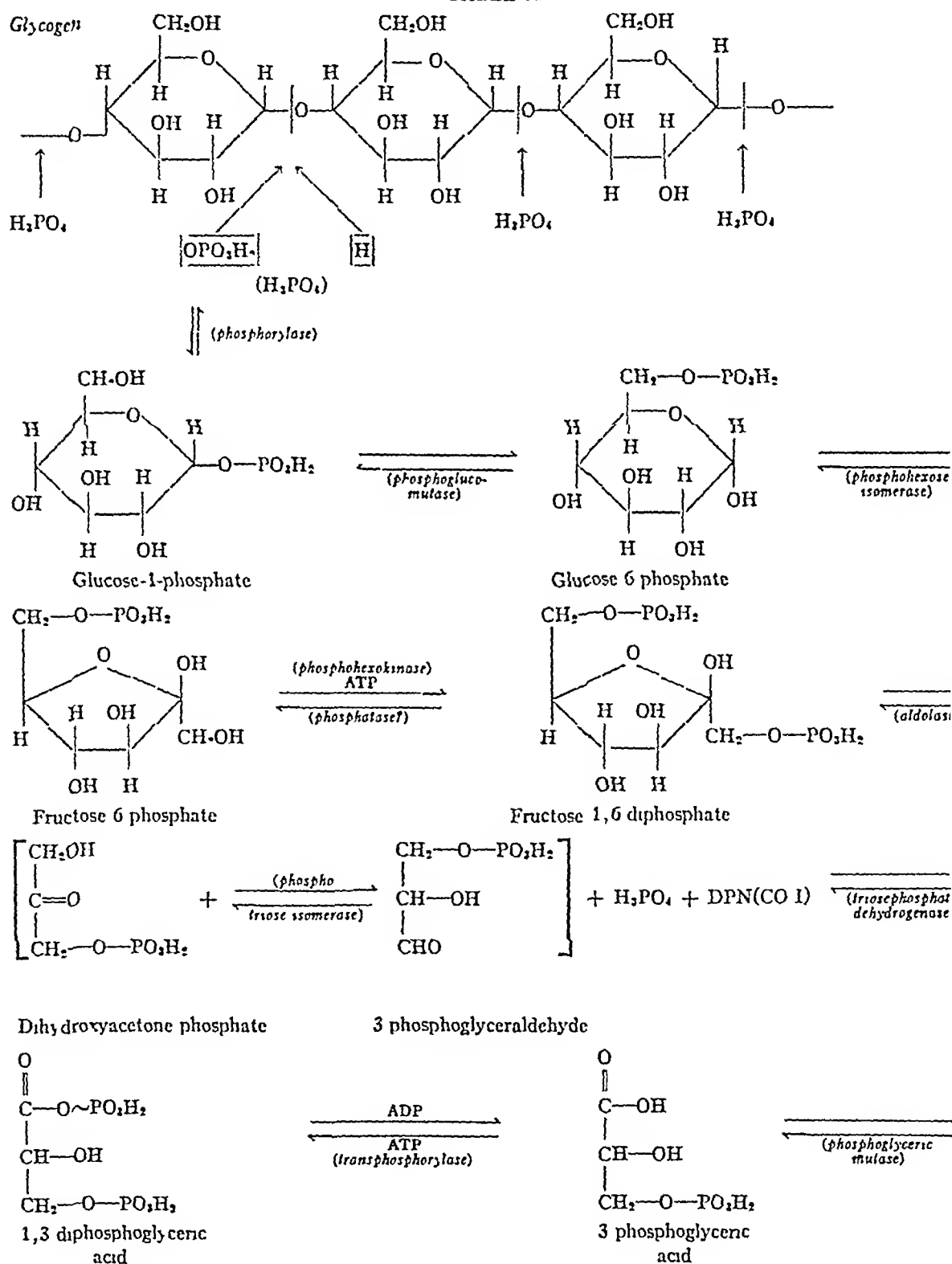
The nucleotide adenosinetriphosphate (also called adenylypyrophosphate or briefly designated ATP) is a compound of adenine, d-ribose and three molecules of orthophosphoric acid. It possesses two energy-rich phosphate bonds, indicated by the symbol \sim in the formula below. The free energy value of each of these bonds is about 10,000 calories per gram mol of phosphate liberated.

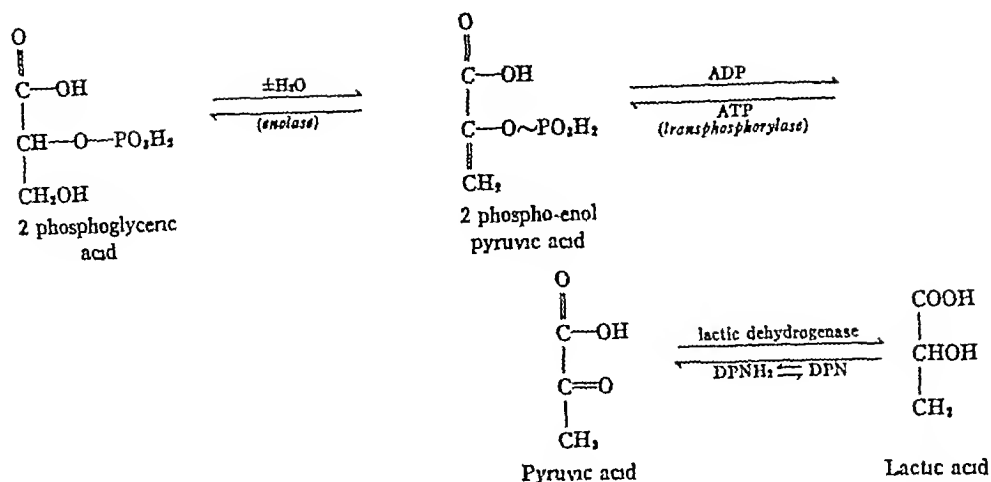


The terminal phosphate group is split off by the action of the enzyme *adenosinetriphosphatase* (ATPase), or the phosphate bond energy may be transferred to other compounds in the course of

glycolysis, ADP is formed. Phosphate is given up, for example, in the conversion of fructose-1-phosphate to fructose-1,6-diphosphate. By reversal of the Lohmann reaction phosphate

TABLE 71





also transferred via ATP to creatine from 1,3-diphosphoglyceric acid in its conversion to 3-phosphoglyceric acid, ADP and phosphocreatine are formed. Energy for the resynthesis of ATP is derived from two main sources, namely, from the Lohmann reaction, $\text{CP} + \text{ADP} \rightleftharpoons \text{C} + \text{ATP}$ and from glycolysis. The first serves at the beginning of contraction or throughout a short contraction, the second, in the subsequent course of a sustained contraction. A third mechanism may come into play but only as a last resort. That is, after the first two sources of energy have been exhausted. This is the formation of ATP and adenosinemonophosphate or adenylic acid (AA) from two molecules of ADP, a reaction which is catalyzed by *myokinase* present in the muscle. Thus,



The ATP/ADP system thus furnishes the immediate energy for contraction, it serves by donating or accepting phosphate to shuttle the latter back and forth, and through the energy-rich bonds, to raise the energy of other compounds to higher levels. The energy of creatinephosphate cannot be transferred directly, but only through ATP.

The transference of the accumulated chemical energy in ATP to the muscle fiber and its transformation to mechanical energy

When adenosinetriphosphate is split the energy

O

of one of its pyrophosphate bonds ($\text{O} \sim \text{P} \quad \text{OH}$)

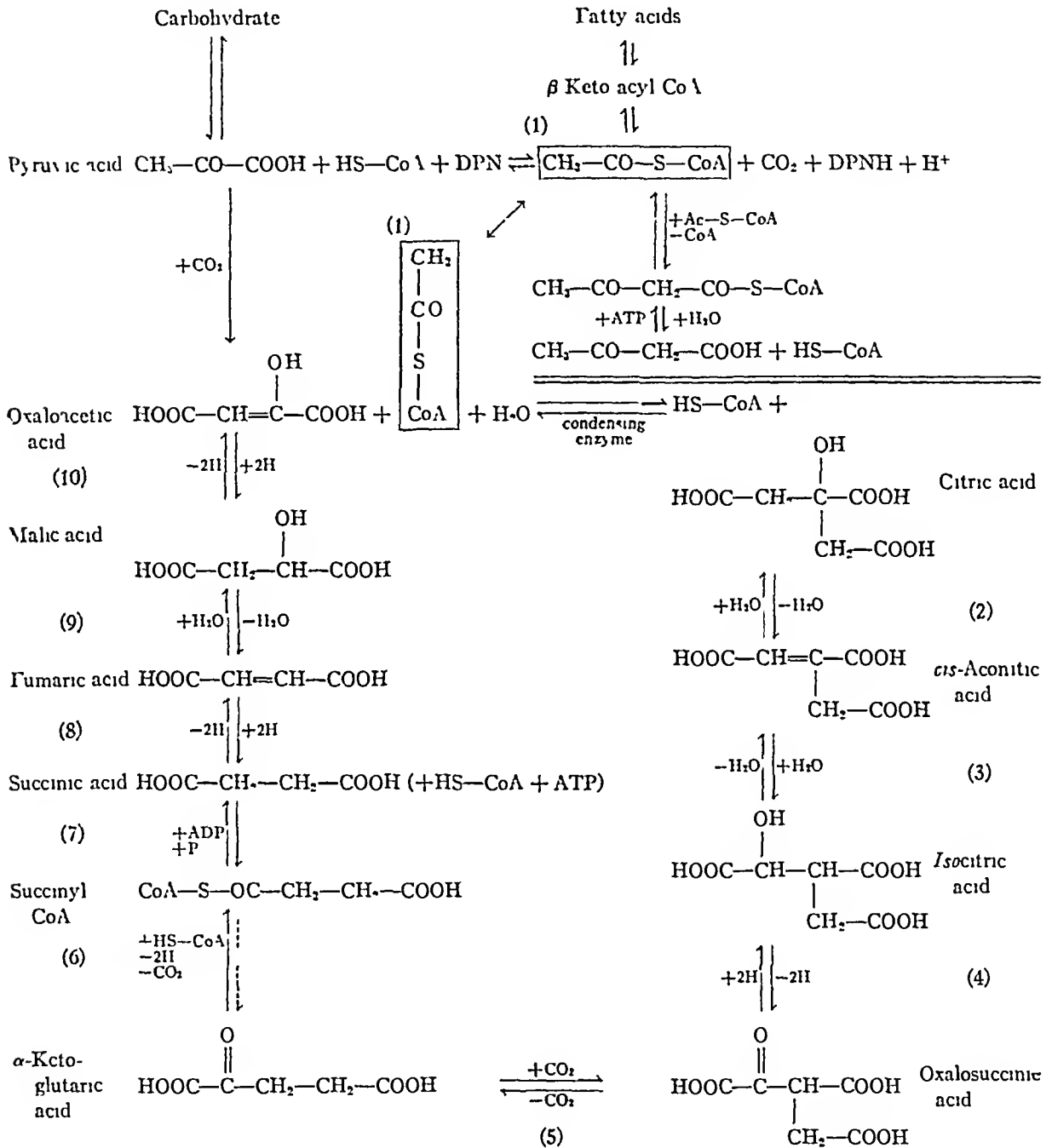
OH

is transformed in the muscle fibers to mechanical energy for the performance of muscular work. After ATP breaks down its energy store is replenished by resynthesis from ADP and phosphate derived from phosphocreatine, as well as from glycolysis. Thus, the ATP/ADP system serves as a "transmission line" by which the phosphate-bond energy of phosphocreatine and glycolytic processes is conveyed to the muscle.

The cleavage of the terminal phosphate group from the nucleotide and the production, as a consequence, of ADP is catalyzed by adenosinetriphosphatase (ATPase). With the discovery by Engelhardt and Lyubimova that myosin, the chief protein of muscle and long recognized as the contractile element, possessed ATPase activity, the chemical reactions were brought into intimate association with the mechanical effects, i.e., with the actual contraction of the muscle. When we come to the muscle itself the energy transformation may be compared, up to a point, with that occurring in an internal combustion engine, in the muscle ATP, being the immediate source of the energy for motion, corresponds to the explosive mixture, and myosin to the piston (Engelhardt), but here the analogy ends for in the muscle machine the myosin "piston" constitutes the ignition system (ATPase activity) as well. The muscle may be compared more closely to an imaginary heat engine in which the piston itself is heated to a temperature where it ignites the gas mixture and the products of the combustion then react upon the piston and produce movement by altering its physical properties. But muscle is not a heat engine, it does not depend upon the pressure of heated molecules for its

TABLE 72

The citric acid cycle in relation to the metabolism of carbohydrate and fatty acids
(Kindness of Professor A M Wynne)



(1) Acetyl Co-enzyme A (Co A)

(2) Aconitase

(3) Aconitase

(4) Isocitric dehydrogenase + TPN

(5) Oxalosuccinic decarboxylase

(6) α -Ketoglutaric dehydrogenase + DPN

(7) Succinyl-CoA phosphokinase

(8) Succinic dehydrogenase

(9) Fumarase

(10) Malic dehydrogenase + DPN

motive power, but upon the energy released through inter- and intramolecular transfer of chemical groups

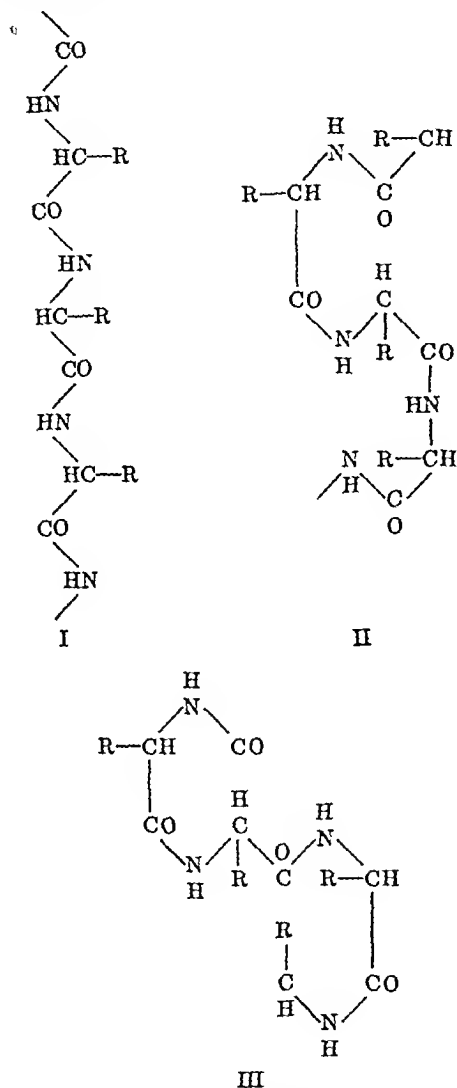


Fig 521 Illustrating the suggested unfolded and folded states of the polypeptid chains in myosin. I, extended (unfolded) state, II, contracted state, III, hypercontracted state R, indicates side chain (After Astbury with minor additions)

Myosin has been obtained by Szent Györgyi in the form of minute needle shaped crystals. Upon repeated recrystallization no loss in ATPase activity, but rather an increase, occurs. From this fact and the following observations, myosin and ATPase are generally assumed to be identical, though absolute proof has not been secured. They

are at least very closely associated. (a) Crude preparations of myosin contain several other enzymes, but these can be removed completely by successive precipitations. (b) The thermolability of myosin and of ATPase are closely similar, the protein undergoing denaturation, and enzyme activity is abolished at about the same temperature, which is relatively quite low. (c) Electrophoretic studies show that myosin possesses a high degree of homogeneity, the large rapidly moving fraction which constitutes over 90 per cent of the protein of the preparation has been found to have double the ATPase activity of the smaller slowly moving fraction which possesses a low protein content. (d) An aqueous (myosin free) extract of muscle has only slight ATPase activity.

Only one phosphate group is split from adenosinetriphosphate by myosin. ADP is not attacked. It shows a high degree of specificity in this regard. Inosinetriphosphate and inorganic triphosphate are split in a similar manner. Calcium ions activate ATPase, whereas magnesium, copper and silver are inhibitory, the enzyme activity is unaffected by cyanide.

It is generally agreed that the immediate and fundamental chemical event in the contraction of muscle is the splitting of ATP by ATPase. But how does this reaction bring about the physical changes in the muscle—the shortening and thickening of the fibers which we call contraction? The researches of Astbury indicate that the polypeptide chains of myosin in the relaxed state of the muscles have a partially folded configuration, but become more fully folded during contraction as shown in figure 521. Thus, like the flexible sides of a concertina, the chains which run in approximately parallel bundles in the micellae shorten or lengthen. Astbury and his associates have compared the X-ray diffraction patterns of muscle and strips of myosin films spread and dried on glass with those previously observed and described for keratin, the fibrous protein of hair, feathers etc. In natural keratin as well as in myosin films and in muscle, the polypeptide chains are partially folded. The natural hair protein, however, differs from the muscle protein in that while it shows reversible extensibility to a high degree due to unfolding and refolding of the polypeptide chains, little shortening from the original length is possible. This is because the parallel chains are cross connected by disulfide ($-\text{S}-\text{S}-$) bridges between the two halves of cystine molecules. When the cross chains

are broken by heat the keratin fiber shortens and the X-ray diffraction pattern resembles closely that of muscle or of a myosin film

Since ATP is present in the muscle during rest and ATPase is closely associated with myosin, the next question which arises is the manner in which a reaction between the two is prevented at this time, and what brings about their interaction when the muscle receives the impulse to contract. It has been suggested that though ATP is bound to myosin in the relaxed state of the muscle, myosin possesses no ATPase activity at this time since calcium ions are not available, and that the impulse liberates Ca^{++} (necessary for ATPase activation) from a compound in which it is bound with protein. A more likely explanation according to Engelhardt is that ATP, instead of being in close association with myosin, is conjugated with some other protein during the resting state. Thus, myosin and another protein are seen as competing for ATP, the latter being successful in the relaxed muscle, and myosin in the stimulated muscle

A protein which appears to be an integral part of the contractile mechanism other than myosin was isolated from muscle by Straub in 1942. This discovery, while it has led to fresh facts being unearthed and is probably of fundamental importance, appears so far to have complicated the picture rather than to have delineated it more clearly. This protein, called *actin* by Straub, has a relatively low molecular weight (55,000–60,000). It constitutes from 12 to 15 per cent of the total protein in muscle. According to Szent Györgi and his associates it is closely associated in the muscle fiber with myosin to form a complex which they call *actomyosin*. Actomyosin can be obtained by prolonged extraction of muscle or by mixing myosin and actin together *in vitro*. The complex, as extracted from muscle, contains 2 parts of actin to 5 parts of myosin. When a concentrated solution of actomyosin in KCl is injected from a syringe into a large body of distilled water a fine thread is formed. This artificial muscle fiber, when immersed in boiled muscle juice, contracts vigorously, becoming both shorter and somewhat thicker. The constituents of the muscle juice responsible for the "contraction" are ATP, KCl and magnesium, for a solution containing these substances in suitable proportions has an effect similar to that of muscle juice. The "contraction" of the actomyosin fiber is in the nature of a rapid and superlative shrinkage

—an extreme syneresis. The water content of the shrunken fiber is only about 50 per cent as compared with 90 per cent in the relaxed fiber. The "contraction" process is reversible, increasing the salt concentration of the solution causes relaxation of the thread to nearly its previous size and form. In order to account for the thickening of the thread as well as its shortening, Szent Györgi proposes that the submicroscopic particles of the two proteins are associated in such a way, namely, short rod-like particles of myosin applied lengthwise to the longer thread-like particles of actin, that shortening of the former causes the conjoined particles to assume a circular form (fig. 52.2)

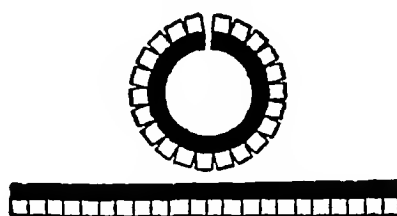


FIG. 52.2 Model composed of a rubber tube slightly stretched to which is attached, prior to stretching, a wooden rod incompletely divided into blocks. When the rubber tube is released, the model must curl up into the shape shown in the upper drawing. The model represents the actomyosin mycel, the rubber tube the myosin, and the wooden rod the actin (Szent Györgyi)

SUMMARY OF THE CHEMICAL CHANGES AND ENERGY TRANSFORMATIONS ASSOCIATED WITH MUSCULAR CONTRACTION

The series of chemical changes through which energy is liberated for muscular contraction may now be given in their natural sequence. The first reaction which can be detected is the cleavage of the terminal phosphate group from ATP. This is catalyzed by ATPase. ADP is produced. Simultaneously, the released phosphate is engaged in the phosphorylation of glycogen, adenylic acid acting in some unknown way as a coenzyme. ATP is resynthesized through phosphate donated by phosphocreatine (Lohmann reaction). The breakdown of glycogen to lactic acid provides the energy for the resynthesis of phosphocreatine and later, as contraction proceeds, for the resynthesis of ATP, in which phosphate-bond energy is stored. These reactions can occur in the absence of oxygen. Upon the admission of oxygen, about $\frac{1}{3}$ of the lactic acid produced is oxidized, this yields the energy for the resynthesis of the remaining $\frac{2}{3}$ to glycogen. When the oxygen supply is adequate little or no lactic acid is formed. In the absence

of oxygen the accumulation of lactic acid slows and then arrests the glycogen \rightarrow lactic acid breakdown and, as a consequence, phosphocreatine and ATP resynthesis is prevented

ATP \rightarrow phosphonic acid + ADP	Energy for contraction (12,000 cal)
CP + ADP \rightarrow creatine + ATP	Energy for resynthesis of ATP (10,000 cal)
Glycogen to lactic acid	Energy for resynthesis of CP and ATP (29,000 cal)
Oxidation of lactic acid	Energy for resynthesis of glycogen (325,000 cal)

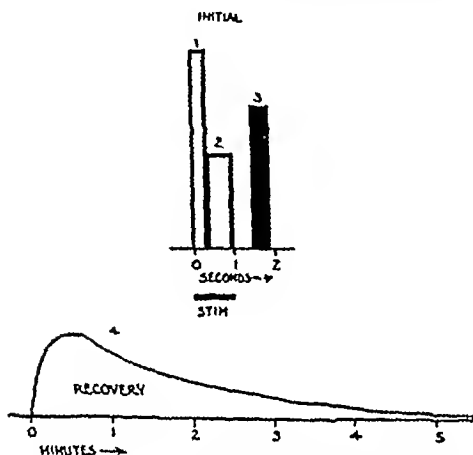


FIG 52.3 Showing the stage of heat production of muscle during and following a short tetanic contraction. The first three stages represent the initial heat 1, the contraction heat, 2, the heat produced during the maintenance of contraction, 3, the relaxation heat. The fourth stage (4) is the recovery heat. (From Evans, *Recent Advances in Physiology*)

HEAT PRODUCTION IN MUSCLE

In the case of a muscle contracting isometrically, i.e., without shortening, all the energy which it expends appears as heat. If, on the other hand, the muscle is allowed to shorten and lift a weight, from 20 to 25 per cent of the total energy expenditure on the average, and under optimal conditions 30 per cent, appears as mechanical work. The efficiency of the muscular machine,

$$\frac{\text{mechanical work performed}}{\text{energy expenditure over the basal level (resting state)}}$$

is therefore much higher than that of the steam engine (7 to 20 per cent) and is comparable to that of the best types of gas engine (25 to 30 per cent).

The heat production of a muscle contracting isometrically in nitrogen shows four phases

- (1) A large outburst of heat at the commencement of the contraction
- (2) A sustained heat production during contraction (tetanus)
- (3) A small outburst during relaxation—relaxation heat
- (4) A small amount of heat produced after contraction and relaxation are over. This so called *delayed anaerobic* heat may continue to be produced for some considerable time

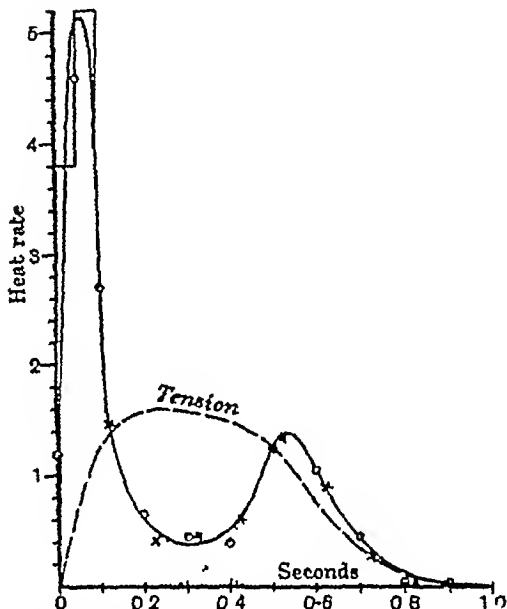


FIG 52.4 Curve of initial heat developed in a single twitch at 0°C. The dotted line starting at 0.25 second, when the tension begins to fall, suggests the correct start for the heat due to mechanical relaxation. (After Hartree)

The first three of these together are called the *initial heat* (figs 52.3 and 52.4). In a single twitch (2) is not evident. The first outburst of heat and, in the case of a tetanus, the intermediary or sustained heat constitute the waste heat of chemical reactions. The outburst during the relaxation or third phase is caused by the degradation to heat of the energy exhibited as tension during contraction; it amounts to about 35 per cent of the entire initial heat. This initial heat is also produced when a muscle contracts in oxygen but it is evolved in the anaerobic phase, i.e., it is independent of oxidative processes.

The total heat production of a muscle contracting in oxygen is over 2.2 times greater than that of one contracting in nitrogen, for there is the

added heat of oxidative processes—the *oxidative heat*. The delayed anaerobic heat plus the oxidative heat is referred to as the *recovery heat*. Thus the total heat evolved during normal contraction is made up as follows

nitrogen but its recovery phase is postponed until oxygen is re-admitted. During strenuous exercise the muscles of the intact animal behave similarly. While working strenuously the circulatory and respiratory systems are incapable of supplying

Initial heat

	<i>arbitrary units</i>	
(1) Contraction heat	0 65	
(2) Relaxation heat	0 35	
		Total initial heat 1 0

Recovery heat

(1) Delayed anaerobic heat, i.e., heat of lactic acid production over that of phosphagen synthesis	0 08	
(2) Oxidative heat, i.e., heat of oxidative processes over that absorbed in resynthesis of glycogen	1 16	
		Total recovery heat 1 24
		<hr/> Total heat 2 24

In a muscle poisoned with iodoacetic acid and contracting in nitrogen, the different phases of the initial heat and its total amount do not differ from those occurring in a normal muscle although lactic acid production has been abolished. In such a case the initial heat production is apparently due mainly to the breakdown of phosphocreatine and adenylypyrophosphate—an explosive liberation of heat. In the normal muscle, on the other hand, lactic acid production and its neutralization by the muscle proteins causes a pronounced evolution of heat (exothermic reaction). Why then are the initial heats the same in both instances? Presumably, in the case of the normal muscle, the heat of lactic acid production and neutralization are masked, being absorbed in the synthesis of phosphocreatine and adenylypyrophosphate (endothermic reaction). In either case it is believed that the initial heat represents the balance of heat production over heat absorption.

The recovery heat is also the resultant of exothermic and endothermic reactions—lactic acid oxidation and glycogen resynthesis. There is also the delayed anaerobic heat. It has been mentioned that a proportion of the lactic acid production occurs after the contraction is over, and that a proportion of the phosphocreatine resynthesis also occurs at this time. For a long time no explanation for the anaerobic delayed heat was forthcoming. It now appears that it, like the initial heat, represents a balance between the heat evolved during delayed lactic acid production and that absorbed by phosphagen resynthesis. The heat of recovery therefore is made up of this delayed anaerobic heat plus the balance between lactic acid oxidation and glycogen resynthesis.

adequate amounts of oxygen for the removal (by oxidation and resynthesis to glycogen, of the large quantities of lactic acid produced. Complete recovery must be postponed until the exercise is over, when the accumulated lactic acid is gradually removed. The muscles of the intact animal during strenuous exertion are therefore comparable to the isolated muscle contracting anaerobically. It was shown by A. V. Hill and his associates that an athlete during great muscular effort such as sprinting cannot possibly inhale more than a fraction of the oxygen required. That is, the body works its muscles but does not furnish them with the total oxygen required for the work until some time after this has been completed—it “goes into debt for oxygen” paying up during the recovery period (see fig. 52.5). In a hundred-yard sprint, for example, the oxygen requirement may be over 6 liters. It is obviously impossible to deliver this amount to the muscle in the few seconds in which the race is run. The maximum consumption possible is not more than 4 liters of oxygen per minute. Furthermore, a sprinter can dash 100 yards with the breath held. The great value of the anaerobic phase of muscular contraction is thus revealed. Through the ability of the muscles to contract when deprived of oxygen and to replenish their stores of energy during the phase of oxidative recovery, they are enabled to perform for short periods an amount of work which otherwise would be impossible, that is, were they, as in the case of a motor engine, dependent entirely upon a contemporaneous oxygen supply.

The oxygen debt is determined by measuring the oxygen used during the period of recovery,

MUSCULAR CONTRACTION IN THE INTACT ANIMAL

Oxygen debt

We have seen that an isolated muscle is able to contract when stimulated in an atmosphere of

i.e., from the termination of the exercise to the time when the oxygen consumption has returned to normal, and subtracting from it the quantity of oxygen used during a corresponding resting period. The length of the recovery period may be 80 minutes or more. In very severe exertion the oxygen debt amounts to over 10 liters (or about 0.3 cc. per gram of muscle tissue). The maximum recorded in man is over 18 liters. During less strenuous exercise the discrepancy between lactic acid production and lactic acid removal is less pronounced, and the oxygen debt is correspondingly smaller. In light exercise the lactic acid is removed during the work—the body “pays as it goes,” and no oxygen debt is incurred. This is called the *steady state*. In other words, anaerobic and aerobic processes are balanced. The average man cannot maintain the steady state unless the oxygen requirement of the work does not exceed

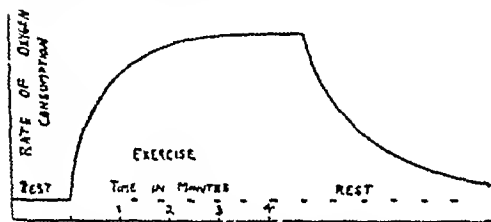
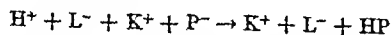


FIG. 52.5 Illustrating the increase in O_2 consumption above the resting level following exercise—“oxygen debt” (From Hill, *Muscular Movement in Man*)

about 2 liters per minute. After severe exercise the normal level of lactic acid in the blood may not be reached until an hour or more after the exercise has ceased.

Lactic acid production during exercise

The lactic acid produced in a short bout of strenuous exercise may amount to as much as 3 grams per second, and its concentration in the blood rise as high as 0.2 per cent. The lactic acid, though buffered by the muscle protein thus

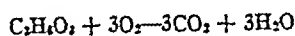


lactic acid	potassium proteinate	potassium lactate	acid pro- teinate undisso- ciated
----------------	-------------------------	----------------------	--

and to a less extent by phosphates and bicarbonates, causes a sufficient change in blood reaction to stimulate powerfully the respiratory center. Large amounts of carbon dioxide are “blown off” from the lungs.

Hill and Lupton have taken the excess oxygen con-

sumed during recovery as a basis for calculating the quantity of lactic acid present in the body at the end of exercise. They assumed that the recovery oxygen was used entirely in the oxidative removal of lactic acid, and that the quantity of lactic acid removed by oxidation was, as in the case of isolated muscle, only about one fifth of the total quantity which disappeared. Thus,



That is, for every 3 molecules of oxygen consumed during recovery 1 molecule of lactic acid has been oxidized, and for every molecule of lactic acid oxidized 4 have disappeared through synthesis to glycogen. Three gram molecules of oxygen consumed therefore represent the disappearance of 5 gram molecules ($90 \times 5 = 450$ grams of lactic acid). Or, 1 gram molecule (22.4 liters) of oxygen represents the disappearance of $\frac{1}{3}$ gram molecules (150 grams) of lactic acid. Each liter of oxygen consumed during recovery would therefore indicate the disappearance of $(150/22.4 = 6.7)$ grams, approximately, of lactic acid.

Non-muscular tissues, liver, heart, brain and muscles not actually engaged in the exercise, share in the removal of any lactic acid carried to them in the general circulation. Barr and Himwich showed this in animal experiments. From the results of their experiments upon rabbit muscle with intact nerve and blood supply, Sacks and Sacks conclude that, contrary to what is believed to occur in isolated frog muscle, lactic acid is not resynthesized to glycogen in the mammalian muscle during recovery, but diffuses into the circulation to be mainly converted in the liver to glycogen. A smaller part is oxidized by brain, heart and probably by other tissues as well. If produced in large amounts, as in strenuous exercise, an appreciable amount finds its way into the urine.

The fuel of exercise

We have seen that in the case of isolated frog muscle the respiratory quotient is around unity, which indicates that the ultimate source of the energy for the contractile process is mainly, if not entirely, carbohydrate. That is, the energy required to restore the muscle to its pre-contraction state is derived from the oxidation of this food material. Though the glycogen of the isolated muscle is reduced by activity, no diminution of its fat content has been demonstrated. It has also been shown that if an isolated muscle be stimulated while suspended in Ringer's solution so that

the lactic acid as it forms may diffuse away, fatigue sets in only when the glycogen stores have been exhausted (see p 715). If glucose be added to the Ringer's, something like 10,000 twitches can be evoked and a total tension of 6 tons per square centimeter of cross-section of the muscle developed. It cannot, however, be concluded from experiments upon the isolated frog muscle, which for one thing is poorly supplied with oxygen, that the muscles of the intact mammal can use only carbohydrate fuel. Himwich and Rose, for example, by determinations of CO_2 and of O_2 of venous and arterial bloods of intact muscles obtained an average R Q of only 0.80. Attempts to decide the question for the intact body have been made by determinations of the respiratory quo-

immediately following the exercise is followed by one in which the output of the gas is reduced and the R Q is well below the normal level of 0.85— CO_2 is being retained to replenish the bicarbonate stores (fig 52.6). During a period of sufficient length, therefore, one effect (retention) will just balance the other (blowing off) and the quantity of CO_2 eliminated during this time in excess of the output during a pre-exercise period will be the extra quantity actually produced by oxidative processes. This value is used in calculating the R Q of the excess metabolism resulting from the exercise.

In experiments on man involving short periods of strenuous exercise, the R Q of the excess metabolism has been found by several observers to rise above that of the resting period and to reach or exceed unity. Best, Furusawa and Ridout obtained a respiratory quotient of between 1.18 and 1.68 for very arduous

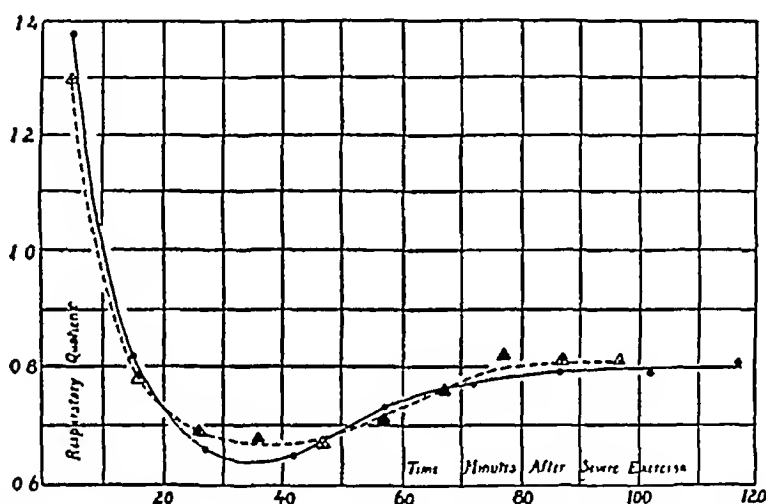


FIG 52.6 Curve of the R Q of the excess metabolism caused by muscular exercise (From Hill, *Muscular Movement in Man*)

tient of the *excess metabolism of exercise* in man. The results are, however, difficult to interpret and no definite conclusion can be drawn from them.

The excess metabolism of exercise is determined from the oxygen consumed (or of the CO_2 eliminated) during the work and recovery periods less the amount of oxygen used (or CO_2 eliminated) during a corresponding period preceding the exercise. The ratio between the excess quantities of the two respiratory gases will, of course, be the respiratory quotient of the excess gaseous exchange. But certain precautions must be taken in order to obtain the true respiratory quotient, i.e., the respiratory quotient of the oxidative processes, for as already mentioned a large proportion of the CO_2 which is eliminated following exercise is not oxidative in origin but is simply gas which has been liberated from chemical combination (p 715). The period of increased CO_2 elimination and high R Q

exercise. For less strenuous work the R Q was around unity and for mild exercise it was considerably lower—little above that for the rest period. The degree of exercise required to raise the quotient to unity varied in different subjects. The very high quotients (above unity) which these observers obtained have since been observed by others, but Gemmill, in more rigidly controlled experiments in which the subject was kept under basal conditions for a period of several hours preceding the exercise, obtained an average respiratory quotient for the excess metabolism of a little less than unity. Determinations of the oxygen consumption and carbon dioxide elimination were made over a recovery (post-exercise) period of 3 hours.

The R Q of the excess metabolism has also been determined by a number of workers upon diets high, respectively, in fat or carbohydrate. Benedict and Cathcart, for example, found an R Q of 0.90 on a carbohydrate-rich diet and one of 0.82 (indicating the

utilization of fat) on the fat-rich diet. They concluded that the fuel of exercise was not exclusively carbohydrate but depended largely upon the previous diet. Krogh and Lindhard decided that the food material oxidized during work was the same as that during rest, and that in either case the relative quantities of fat and carbohydrate oxidized varied with their proportions in the diet. Work, however, was performed less economically (by 10 per cent) upon fat than upon carbohydrate. Others have observed an increase in efficiency upon high carbohydrate diets, though, as a rule, the increase was not as great as that found by Krogh and Lindhard.

Carpenter and Fox carried out two comparable groups of experiments. In one group 50 grams of glucose were given just prior to the commencement of the exercise. In the other group the subjects were fasting. The R Q of the excess metabolism of those which had received the glucose was considerably higher than those of the other group.

Anderson and Lusk obtained in a starving dog working a treadmill respiratory quotients suggestive of the combustion of fat, namely, from 0.71 to 0.73. Also in prolonged severe exercise in man Talbot and associates observed a fall in the R Q following an earlier rise.

Other lines of approach have yielded more important evidence for carbohydrate being the main fuel, though not necessarily the only one, used in muscular work. Other food-stuffs, especially fat can serve under certain circumstances, the proportion of each which undergoes oxidation varying with the amount available and the severity of the exercise. In *short periods of light exercise* it is probable that the energy is derived from the oxidation of materials of the same nature as those which furnish energy during rest. These are small molecules of carbohydrate, fat and protein material already present in the circulation (Carpenter). Such materials are soon exhausted, however, by heavier work, when the glycogen reserves are drawn upon. There is general agreement that in *strenuous exercise* the fuel used by the muscles is mainly if not entirely carbohydrate. If the exercise is long continued, as in a marathon race, hypoglycemia results. The administration of glucose at the commencement of the race prevents the fall in blood sugar and exerts a definitely beneficial effect upon the athlete's performance. The production of large quantities of lactic acid in strenuous muscular exercise also points to carbohydrate material as the source of energy. In short but strenuous bouts of work a rise in blood sugar

up to a concentration of nearly 200 mg per cent has been observed. The mechanism governing this mobilization of the carbohydrate reserves has not been definitely determined. It does not appear to have an emotional basis (liberation of adrenaline or stimulation of the sympathetic). Gemmel suggests that the high blood lactate stimulates the production of glucose by the liver.

In prolonged and exhausting exercise the carbohydrate stores become depleted. Fat and, to some extent, protein are then utilized to drive the muscular machine. Both these materials are used indirectly, i.e., after their conversion to carbohydrate. As the carbohydrate stores are used up physical exhaustion ensues. Dogs can be worked for 17 hours or so before they are completely fatigued if fed carbohydrate, but become exhausted in less than 5 hours if worked without food (Dill and associates). The fatigued animals are capable of further work if then given carbohydrate.

The use of protein as fuel has been a subject of controversy for a number of years. It has been common teaching that this food material, except in minimal quantities when fat and carbohydrate are unavailable is not a source of energy for muscular work, but served merely to repair tissue "wear and tear". This conclusion was arrived at chiefly from studies of the nitrogen excretion. Ordinary exercise, for example, does not increase appreciably the output of nitrogen in the urine nor does it increase the non protein nitrogen of the blood. More strenuous work causes a slight rise in the blood non protein nitrogen and a moderate increase in the urinary nitrogen in man and in animals. In work experiments upon fasting dogs it has been calculated that at the most not more than 7 per cent of the energy required for the exercise could have been derived from protein, the great part of the energy had apparently been obtained from the combustion of fat. In prolonged starvation after exhaustion of the stores of carbohydrate and fat, protein (carbon part of the amino acids) must then, serve as the sole source of energy.

It has been pointed out by Cathcart, however, that the nitrogen excretion during short periods of exercise may not be a true criterion of protein metabolism since the nitrogen released in the breakdown of protein may be utilized in synthetic processes and consequently not appear in the urine. Or, muscle protein may be catabolized and its nitrogen excreted, yet if, as is quite possible, an equivalent amount of nitrogen derived from the food were diverted to the muscles to replace that which had been lost, the total nitrogen excretion would remain unchanged. This observer also con-

siders that, in the long run, muscular work exerts a very definite influence upon protein metabolism and cites the familiar observation that persons engaged in heavy muscular work demand a diet rich in protein, particularly meat. It has also been shown by several investigators that a retention of nitrogen occurs during a period of training—apparently for the manufacture of muscular tissue

Such views are in harmony with modern conceptions of protein metabolism (p 631) Since catabolism and synthesis may go hand in hand the total quantity of nitrogen excreted gives no indication of the interchanges which are taking place between food and tissue nitrogen It is scarcely reasonable to assume that after deamination the non-protein portions of the amino acids can not serve as a source of oxidative energy A carnivorous animal, for example, can subsist upon a diet composed almost exclusively of protein Canzanelli and Rapport also draw attention to the irrelevancy of the nitrogen excretion and insist that it can give no information respecting the non-

nitrogenous part of the catabolized protein From a study of the respiratory quotient of the excess metabolism of exercise performed by a dog on a high protein diet, they conclude that oxidative energy for muscular exercise can be supplied quite as readily by protein as by carbohydrate or fat.

The question whether or not *alcohol* can furnish energy for muscular work has been investigated repeatedly The results obtained by different workers are not entirely in agreement Some have reported that muscular exercise hastened the disappearance of alcohol from the blood, presumably by increasing combustion (see Mellanby) On the other hand, Carpenter and his associates in a recent careful study did not find that work exerted such an effect They conclude that alcohol disappears from the human body at a uniform rate whether the subject is at rest or performing muscular exercise Nor did muscular work alter the concentration of alcohol in the expired air (i e the amount eliminated per liter of CO₂ remained unchanged), in the urine or in the blood

THE BODY TEMPERATURE HEAT BALANCE

The normal body temperature, recorded from the mouth, is usually given as 98.6°F (37.0°C). The rectal temperature averages $\frac{1}{10}$ of a degree F higher than this and the axillary temperature about $\frac{1}{2}$ a degree F lower. No absolute figure can be given, for there is a wide variation between individuals. The oral temperatures of a large group of normal persons range from 96.6° to 100.0°F (35.8–37.8°C) the average lying around 98.4°F (36.9°C). Ivy, in a study of nearly 300 medical students, obtained a mean oral temperature of 98.1°F. Variations in the body temperature also occur in the same individual throughout the day—a difference of 0.5° or even 1.0°F occurring between the maximum in the late afternoon or early evening, and the minimum at about 4 or 5 o'clock in the morning. In night workers the times of the maximum and minimum temperatures may be reversed. The temperature of the internal organs is higher by from 2° to 3°F or more than the temperature of the skin. The temperature of the liver, for example, is about 100°F whereas that of the skin covered with clothes is from 85° to 93°F. The temperature of the bare skin varies widely, of course, with the environmental temperature. The influence of the latter upon the temperature of the covered skin will depend upon the heat-insulating properties of the clothing, air movement (breeze, wind), atmospheric moisture, etc. Strenuous muscular exercise causes a temporary rise in body temperature of 1.0° to 4.0°F or more, a temperature of over 104°F during exercise has been reported (L. Hill).

The heat-regulating mechanisms are not fully developed at birth. The body temperature of the newborn child, though in general the same as that of the adult tends to be irregular and unstable. Spontaneous variations of from 1 to 2 degrees are common during the first year. Excitement or other strong emotion, even in older children, may raise the temperature by as much as 2°F.

THE REGULATION OF BODY TEMPERATURE

Mammals and birds possess efficient mechanisms for maintaining a constant body temperature against extreme changes in environmental temperature. It is a remarkable fact that the tem-

perature of a warm blooded (homoiothermic) animal remains practically unchanged though the surrounding temperature may vary between 0°F or less and 100°F or upwards. On the contrary, the body temperature of a cold-blooded (poikilothermic) animal such as the frog, turtle, etc. is practically that of its environment (fig. 53.1).

The body temperature of the homoiothermic animal represents the balance struck by the heat produced in the tissues (and heat acquired in warmed food) and the heat lost to the environment. The body also absorbs heat radiated from surrounding objects with temperatures higher than its own, and from direct or reflected sunshine, a stove or an open fire. Heat production is the result of chemical reactions and is therefore spoken of as the *chemical regulation* of body temperature. Heat loss is dependent upon physical (and physiological) factors—*physical regulation*.

The body's heat-regulating mechanism is supposed to be in abeyance below an internal temperature of 75°F, the body then gaining or losing heat like an inanimate object.

PHYSICAL REGULATION—HEAT LOSS

Heat is lost from the body through

- (a) *Radiation, convection and conduction*
- (b) *Evaporation of water from the lungs and skin*
- (c) *Raising the inspired air to body temperature*
- (d) *Urine and feces*

Under the ordinary conditions of every day life over 95 per cent of the total heat loss occurs through (a) and (b). The heat lost in raising the temperature of the inspired air to body temperature (c) will, of course, vary with the air temperature, but at ordinary room temperatures it does not amount to more than 2 or 3 per cent of the total. The air is a very poor conductor, so in terrestrial animals conduction plays a very minor role except under special circumstances, as when the body is in contact with a cool object, cold ground or immersed in water. Radiation is responsible for about 50 per cent of the total heat loss and convection for about 15 per cent (see table below). The heat lost in the urine and feces accounts for only 2 per cent or less of the total heat loss.

The total quantity of heat lost in twenty-four hours must, of course, just equal the amount produced, otherwise the body temperature would rise or fall. The heat production of an average man doing light work is about 3,000 Calories. The proportions of this which are dissipated through the various channels at ordinary room temperature are given in the following table, in approximate figures

	Calories	Per cent
(a) Radiation, convection and conduction	1950	65
(b) Evaporation of water from skin and lungs, and liberation of CO ₂	900	30
(c) Warming inspired air	90	3
(d) Urine and feces (i.e., heat of these excreta over that of the food and water)	60	2
Total daily heat loss	3000	100

Radiation and convection

The loss of heat by these means varies with (a) the air temperature and other environmental conditions, e.g., humidity and air movement, (b) the nature and amount of clothing, and (c) the quantity of heat produced within the body, i.e., with the metabolism.

The rate of cooling of any warm object varies with the temperature of the air and of colder objects in contact with or near it. When a large temperature difference exists between the two, the warm object loses heat rapidly through radiation and convection, the rate of heat loss, however, becoming gradually less as the temperature of the object approaches that of the environment. The dead human body behaves in a manner similar to that of any inanimate object, taking from ten to twenty hours on an average to reach the temperature of its surroundings. In the living body, on the other hand, factors operate to encourage or minimize heat loss, respectively, when the environmental temperature is high or low, or corresponding changes in the body's heat production occur. The factors involved in heat conservation or heat loss are dependent essentially upon reactions of the autonomic nervous system. The following are the principal adjustments which take place in the blood-vascular system: (a) *Redistribution of blood*. The cutaneous vessels dilate or constrict and through the diversion of blood from internal regions of the body to the surface, or from the surface to the internal organs, heat loss is increased

or diminished, respectively. At a room (ambient) temperature of 34°C the quantity of blood circulating through the skin may, according to DuBois, amount to 12 per cent of the cardiac output. These changes may be initiated in one or all of four ways, a change in temperature of the blood supplying the nervous centers, reflexly through centers in the brain and cord in response to changes in skin temperature (stimulation of hot or cold spots, p. 934), through local axon reflexes, and finally through responses of the vessels to direct stimulation by changes in external temperature. (b) *Variations in blood volume* (see p. 25). A rise in temperature causes an increase in blood vol-

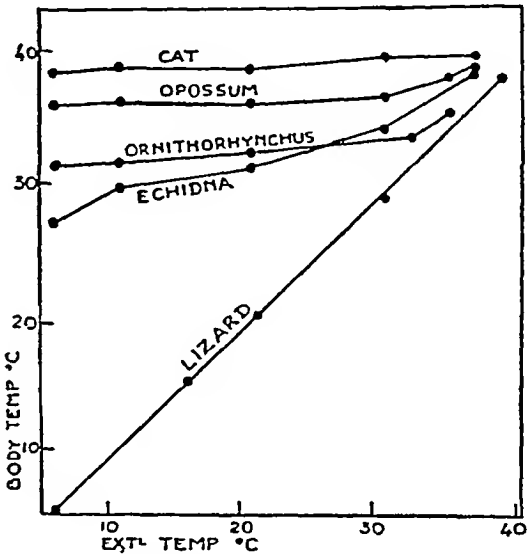


FIG. 53.1 Variation of body temperature of different types of animals by sojourn for two hours in an environment of 5° to 35°C (After Martin)

ume, the blood is diluted by fluid drawn into the circulation from the tissues, chiefly the skin, muscles and liver. Blood is expelled from the spleen (p. 70). At low temperatures the blood volume is reduced, the blood becoming more concentrated as shown by an increase in the percentage of blood solids. These changes in blood volumes are of paramount importance in the regulation of body temperature. Barbour has shown that when a dog, whose spinal cord has been sectioned in the lower cervical region, is immersed in a cold bath, concentration of its blood does not occur and its temperature falls. When a normal animal is exposed to cold in a similar manner, concentration of the blood occurs and the body temperature remains practically un-

changed (fig 53.2) (c) *Increased circulation rate* (p 265)

The epidermis and the subcutaneous tissues when the vessels are constricted are a little more efficient as insulating material than a layer of cork of the same thickness (DuBois) The subcutaneous layer varies considerably in thickness in different persons and is thicker in women than in men This accounts largely, no doubt, for the

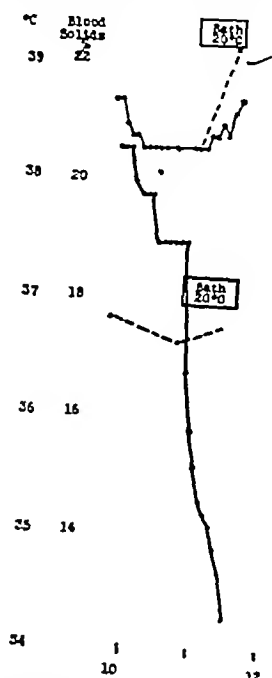


FIG 53.2 Effects of cool (20°C) bath upon dog with and without nervous heat regulation. Upper two curves, normal dog, lower two curves, dog after section of sixth cervical cord segment, continuous lines temperature, broken lines, blood solids. Note that normal dog keeps its temperature from falling by means of blood concentration. Dog with sectioned cord is poikilothermic because blood-concentrating mechanism has broken down (After Barbour)

greater resistance of the former to low temperatures. For the same reason persons who are overweight, owing to the better insulation provided by the excess subcutaneous fat, are better able to withstand the cold than lean persons, but they also become more easily overheated when, as in muscular exercise, heat production is increased.

Convection, i.e., the rate of movement of warm air from the neighborhood of a heated object varies, of course, with the temperature of the atmosphere. The clothed body has a layer of warm

moist air in contact with the skin which tends to become trapped in this situation and in the spaces of the clothing. In the absence of a temperature difference between it and the external air or of some movement to cause mixing, this air will remain practically stagnant, and little or no heat is lost through convection. However, when the atmosphere is cooler, convection currents are set up which mix the air lying against the skin with fresh air. Convection is essentially dependent upon the relative densities of airs at different temperatures, the warmer and lighter air rising, the cooler air falling. Dry air is denser than air possessing a high content of water vapor. One would expect therefore that changes in the humidity of the atmosphere should alter the heat lost by convection but, as a matter of fact, the relative humidity of the external air has little effect upon heat loss through convection. Probably the most important factor influencing heat loss by convection, is air movement, the loss increasing with the square of the wind velocity up to around 60 miles per hour, a wind velocity beyond this exerts little or no further effect.

As already stated, half of the total heat loss is brought about through radiation (p 731). The human skin (of whatever color) within the range of the infra-red to which it is usually exposed, is an almost perfect "black body radiator." That is to say, it radiates nearly all infra-red¹ rays (up to 1 or 2 per cent) or absorbs to the same extent such rays as fall upon it (Hardy and Muschenheim). The radiating surface of the standing human body² is only about 85 per cent of the total surface area, for apposed surfaces, e.g., axillae, inner surfaces of the thighs and upper arms, do not lose heat to the environment by radiation. It is for this reason that one huddles in the cold and spreads out the limbs in the heat. The main factor influencing heat loss through radiation is the temperature of surrounding objects relative to that of the skin. The body, for example, radiates heat to a block of ice but absorbs heat from a hot stove or radiator. It should be remembered that the air intervening between the body and the source of heat is not heated by radiant energy, but only by convection. Another factor, though a very minor one, is the humidity of the atmosphere. Air with a high water vapor content is more opaque to radiant heat than dry

¹ The wave length of the infra red rays emitted by skin at usual temperature (34°C) is 9440 mμ.
² This has been called the "profile area"

air Heat lost through radiation is therefore slightly reduced when the relative humidity is high

Evaporation of water

It is obvious that the nearer the temperature of the environment comes to that of the blood the smaller will be the amount of heat which can be lost by radiation and convection At an air temperature of about 98.6°F heat loss by these means must cease At higher air temperatures than this, the body, were no other means of cooling available, would actually gain heat Through the secretion and evaporation of sweat and the exhalation of water vapor (expired air is practically saturated with moisture) large quantities of heat are lost to the body Its temperature can, for this reason, be maintained constant when the atmosphere (dry) has a temperature about 150°F above that of the blood (see p 734) The heat absorbed in the evaporation of 1 cc of sweat amounts to 0.58 Calorie Even at ordinary room temperatures when there is no obvious perspiration the heat lost through evaporation from the lungs and skin amounts to from 25 to 30 per cent of the total heat loss At higher temperatures the increase in the proportion of heat lost by evaporation of water as compared with that lost by radiation and convection is shown in figure 53.3 It will be seen that evaporation plays little part in heat regulation until the air temperature reaches between 28° and 30°C, the heat loss by this means remaining nearly constant below this level but increasing rapidly above it At a temperature above 35°C evaporation accounts for all or nearly all the heat lost from the body

It is to be remembered that evaporation from the body surface occurs quite independently of sweat secretion, for the skin is not entirely impervious to water, fluid extravasated from the cutaneous capillaries seeps into the epidermis It has been shown in persons in whom sweat glands were absent from birth that some 18 grams of water per square meter of body surface may be lost hourly by evaporation This is about the same as that of a normal man under ordinary conditions, and represents a total daily heat loss of about 450 Cal for a body of average size (surface area 1.8 square meters), which is not far from the normal At high temperatures, or even during mild exercise, however, a person without sweat glands is at a great disadvantage, his body temperature is likely to rise.

The respirations are increased by a rise in air temperature or by a greater heat production, the heat loss through warming the inspired air and the vaporization of water from the lungs, and the liberation of CO₂, is thereby increased Hyperpnea (panting) is the chief means possessed by the dog (in which functioning sweat glands are largely confined to the foot-pads) for increasing the vaporization of water and combating a rise in body temperature

The rate of evaporation of water is influenced inversely by the degree to which the atmosphere is already saturated with moisture, i.e.,

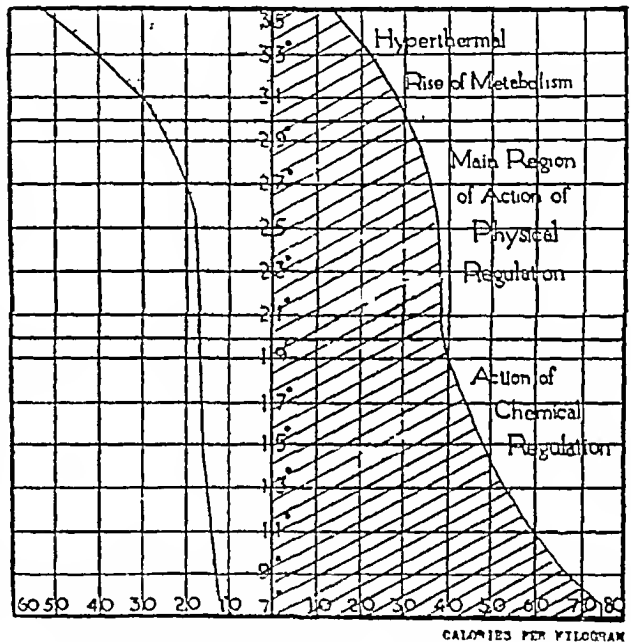


FIG 53.3 Rubner's chart showing the manner of heat loss in the dog at different room temperatures Stippled area, heat loss in calories through evaporation of water, cross-hatched area, heat loss in calories through radiation and convection The distance between opposite points on the curved boundary lines represents the total heat production at a particular temperature (Redrawn and slightly modified from Lusk.)

by its relative humidity³ Sweat which is not evaporated but simply drips from the skin, of course does not increase heat loss For this reason the sweating mechanism for the elimination of heat is badly crippled when the relative humidity is high⁴ We are all familiar with the insufferable

³ It is dependent essentially upon the difference in the vapor pressure at the skin and of the surrounding air

⁴ The relative humidity is defined as the ratio of the weight of water vapor contained in a given volume of air to the weight which the same volume of air would contain when saturated The quantity of water vapor

heat of a hot humid day, and the comparative comfort when the air is simply hot and dry. A man can maintain a normal temperature in an atmosphere of from 240° to 260°F (a temperature that will grill a beefsteak) provided the air is perfectly dry, the ability to sweat profusely is sustained by large droughts of water and evaporation is facilitated by stripping the greater part of the body's surface. On the other hand, a damp atmosphere with a temperature of 120°F causes the body temperature to rise rapidly, and cannot be endured for more than a few minutes. Evaporation and consequently heat loss by this means is greatly hastened by air movement. The layer of air nearly saturated with water vapor lying next the skin is thus replaced by drier air.

SWEATING

Sweat is a weak solution of sodium chloride in water together with urea and small quantities of potassium and lactic acid. It has a specific gravity of from 1.002 to 1.003. Its pH as reported by different observers varies from 4.2 to 7.5. The percentage of sodium chloride varies between 0.2 and 0.5. Muscular exercise increases the salt concentration, which is also higher in sweat secreted by clothed than by naked skin. The quantity of non-protein nitrogen ordinarily excreted in the sweat, per day, is according to Benedict 0.071 gram, on the other hand, if the sweating is copious from 0.5 to 1.0 gram may be eliminated per hour. The actual concentration of nitrogen in the sweat falls, however, when sweating becomes profuse, whereas the concentrations of sodium chloride and potassium rise. However, when acclimatization has been established the percentage of salt diminishes. About 3 grams of salt daily are required after acclimatization to maintain the salt balance. Nevertheless, even after acclimatization, if strenuous work is performed for a long period in a high temperature, and large quantities of water are drunk, depletion of the body's supplies of chloride and a lowered concentration of this element in blood and tissue fluids result.⁵ Severe cramps

which air can hold when saturated increases with the temperature. The relative humidity is expressed as a percentage. Thus, if a sample of air at a certain temperature contains 20 per cent of the water vapor which it is possible for it to contain at that temperature, it is 20 per cent saturated, and so has a relative humidity of 20 per cent.

⁵ A miner performing heavy work may lose a quart of sweat per hour. In a very hot environment arduous work may entail a loss three times this amount.

occur in the muscles of the limbs and abdominal wall ("stoker's" or "miner's cramp"). In order to prevent these effects it is recommended that the thirst be quenched with a weak salt solution (0.2 per cent) instead of with water.

The control of sweat secretion

The sweat glands, which number over two million in a man living in temperate climates, are under the control of the sympathetic nervous system. These glands are, however, anomalous in their responses to sympathetic and parasympathetic drugs, in that they are stimulated by muscarine, pilocarpine and acetylcholine, and inhibited by atropine (ch. 72). According to Coon and Rothman, the action of acetylcholine is two fold,—stimulation of the glands directly through a muscarine like action and through the initiation of axon reflexes (nicotine like action). In man and most animals, they are not excited by adrenalin nor paralyzed by ergotoxine. The usual stimulus to sweat secretion is a rise in blood temperature which exerts its effect in two ways—directly upon the nervous centers, which is of more importance, and reflexly by stimulation of heat receptors in the skin. The sweat response to a rise in temperature is abolished by sectioning the nerves to a part and is therefore not due to direct stimulation of the glands. That a rise in temperature of the centers alone will induce the secretion of sweat has been shown by heating the carotid blood in the cat (whose sweat glands are confined to the paw pads), sweating then occurs though the paws themselves remain cool. The centers may also be stimulated in man by the injection of pituitrin into the lateral ventricle (p. 798). In the initial stages of muscular exercise sweating is apparently due to the discharge of impulses from the motor cortex. It occurs before there is any change in rectal temperature. Later on, the effect of a rise in body temperature comes into play. In a man (indoor clothing) at rest, visible sweating usually commences at an air temperature between 80° and 90°F. Sweating may be induced by the experimental stimulation of regions of the diencephalon (hypothalamus, p. 1026). Spinal centers also exist, since after complete transection of the cord, reflex sweating occurs in the parts of the body below the level of the lesion interruption. Destruction of the sympathetic nerve supply to a part completely abolishes the sweating response to a rise in temperature. The sweat glands, however, still respond to pilocarpine and adrenalin. The former drug, which has

been employed in the past for the purpose of demarcating areas deprived of their sympathetic supply, is of no diagnostic value, for it acts peripherally, i.e., directly upon the gland cells. Sweating is not dependent upon the circulation for it occurs after occlusion of the vessels and can even be induced by stimulation of the nerves in an amputated limb. Though usually associated with cutaneous vasodilation it may occur with constricted vessels—*cold sweat*. This is usually the result of psychic influences, e.g., nervousness, fear, fatigue or mental work. The sweating occurs most noticeably in these instances on the forehead, the palms of the hands, and the soles of the feet, which do not, as a rule give a pronounced response to heat. In many persons reflex sweating, confined to the face and neck is induced by eating spicy or appetizing food (gustatory sweating), or sometimes sweating over a remote part, e.g., the knee, occurs. Faradic stimulation of the human skin over the forearm induces local sweating due apparently to direct stimulation of the glands, for it occurs after section and degeneration of the nerve supply.

The few observations that have been made upon the secretion pressure of sweat indicate that it is high—250 mm of mercury or more. Sweat is therefore a true secretion and not simply a filtrate. The secretion rate may be enormous, amounting to a liter or more per hour, and may be increased some 80 times over the normal by immersing the body in a bath at 108°F. At ordinary room temperatures the sweat evaporates as quickly as formed, so that there is no apparent secretion. The loss of sweat in this way together with the evaporation of water from the lungs and from the surface of the body independently of the sweat glands is called *insensible perspiration*. Its amount varies directly with the basal metabolism.

THE CHEMICAL REGULATION OF BODY TEMPERATURE—HEAT PRODUCTION

The several factors which stimulate the chemical processes of the body and so increase the heat production have been dealt with in chapter 46. An account of the manner in which chemical and physical factors interact to maintain a constant body temperature remains to be given.

A low environmental temperature is a potent influence in stimulating heat production. At air temperatures below about 28°C the body (nude) loses heat rapidly. Within the temperature range between 28° and 30° or 31°C the naked male body is able quite easily to maintain the balance between heat loss and heat production. There is neither sweating nor shivering and a male sub-

ject feels comfortable. This range of temperature is therefore called the *comfort zone*. For reasons given below the comfort zone is broader (27° to 32° or 33°C) for women. The external temperature (about 23°C) below which heat production must be increased in order to maintain a normal body temperature is sometimes called the *critical temperature*.

Below the critical temperature radiation of heat from the body increases progressively with falling air temperature, but heat loss by convection and vaporization shows little change. The naked body at a temperature below the critical level (28°C) loses more heat than it can produce in the basal state, and at about 23°C, a chill (shivering) occurs. Heat production is thus increased in an effort to raise the body temperature to the normal level. In the human subject heat production is not increased until the onset of the chill and in nude men under basal conditions the metabolism remains constant within the range of air temperature from 35° to 23°C.⁶ It has been shown by Hardy and DuBois that this is not true for women. They show a *reduced* heat production of from 14 to 20 per cent at temperatures between 30° and 32°C. Also, owing to the greater insulation afforded by the thicker layer of subcutaneous fat the heat loss of the female body in a cold environment is some 10 per cent less than that of men. Thus women have a more efficient thermo-regulating mechanism than have men, being better able to maintain the heat balance at lower temperatures without shivering, and also to be more comfortable at higher temperatures.

The critical temperature and the temperature of comfort will vary of course with the amount and nature of the clothing. The cooling effect of water is some 20 times greater than that of air—a cold bath at 40°F increasing the heat production some 5 times above the basal level.⁷ Few men can survive for long in very cold water—below 4°C (about 40°F) which is around the winter temperature of the sea in temperate latitudes. Heat production becomes depressed after immersion for 20

⁶ Rubner believed that chemical regulation at low environmental temperature involved some factor other than an obvious increase of muscular activity. In animals, adrenaline is liberated during short exposures to cold and stimulation of the thyroid for prolonged periods has been demonstrated (p. 826), but chemical regulation of this nature has not been shown for man.

⁷ In cold-blooded animals, which possess no chemical regulation, the metabolism as measured by the carbon dioxide output *falls* with the environmental temperature.

minutes or so, and the body temperature falls. In an experiment in Germany during the last war, one man survived a temperature of 5.5°C (42°F) for three hours. His body temperature was then 77.3°F.

It will be seen from figure 53.4 that heat loss increases both above and below the critical temperature. At the lower temperatures heat is lost mainly by radiation and convection, at the higher temperatures mainly by vaporization. It will also be noticed that at the lower temperatures the skin temperatures follow a straight line, but that the curve commences to flatten out at around 30°C.

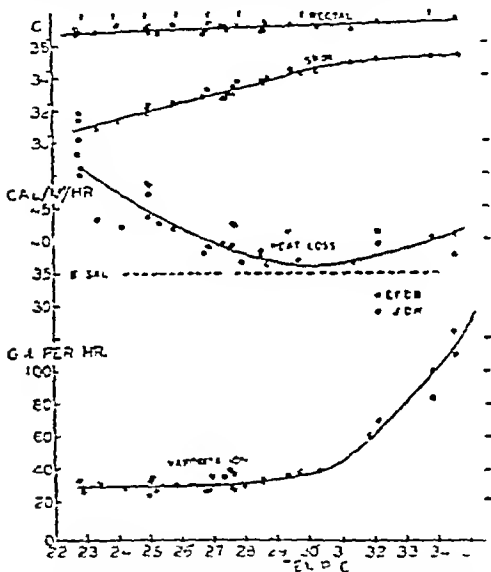


FIG. 53.4. Showing effect of a rising air temperature on rectal and skin temperatures, heat loss, and vaporization. (Modified from DuBois.)

This is attributed to dilatation of the cutaneous vessels and the flooding of the skin with blood.

The muscular tissues (particularly of the extremities) and the liver, wherein numerous chemical reactions are carried out, are the main sources of the body's heat. The rise in metabolism which results from a fall in atmospheric temperature is effected through an increase in tone of the skeletal muscles and in some instances by fine involuntary contractions, e.g., shivering and chattering of the teeth. Contractions of the smooth muscle of the skin, giving rise to "goose flesh" also contribute. In some mammals and in birds, the contraction of the cutaneous muscles also curtails heat loss through ruffling of the hair or feathers. After the skeletal muscles are paralyzed by curare, an ani-

mal loses the power to maintain a normal body temperature in a cold environment. Its physical mechanisms of control are, however, intact so that it can resist high air temperatures. Isolation of the muscles from control by section of the cord in the lower cervical region (C6) also abolishes the chemical regulation. This operation abolishes physical regulation as well, for, as already mentioned, concentration of the blood does not occur, vasoconstrictor reactions are reduced and the body temperature falls when the animal is exposed to a low temperature (fig. 53.2). The rise in body temperature in strenuous muscular effort is apparently due, not to any failure of the heat-dissipating mechanisms, but rather to the "thermostatic" control being set at a higher level, for the temperature rise is about the same whether the exercise is performed at an air temperature of 3° or of 23°C. There is a great increase in the heat loss after the

TABLE 73

ATMOSPHERIC TEMPERATURE °C	HEAT PRODUCTION IN CALORIES PER KILOGRAM OF BODY WEIGHT		
	Starved	100 grams of meat fed	Lowest
4.2	128	133	4
14.5	101	111	9
22.1	71	101	43
30.7	62	117	87

exercise, the heat retained during the exertion being then dissipated.

Food, especially protein through its specific dynamic action, is an important factor in the chemical regulation of body temperature. At high environmental temperatures the specific dynamic action of food acts counter to the physical mechanisms which hasten heat loss. For this reason a low protein diet is more suitable in hot weather. At very low atmospheric temperatures, on the contrary, the specific dynamic effect of food is almost completely masked, since it simply replaces the environmental effect (cold) upon heat production. That is to say, the neuromuscular mechanisms called into play to increase heat production at low temperature are less necessary since the food itself raises the level of metabolism. Protein food in cold climates is therefore a valuable aid to the chemical regulation of body temperature.

The effect of protein ingestion upon the heat production of a dog at different temperatures is shown in table 73 (Lusk).

CONTROLLING CENTERS

Section of the neuraxis through the mid-brain at the level of the superior colliculi, or at any level posterior to this and anterior to the lower cervical cord, renders an animal poikilothermic (fig 53 5). Section of the cord in the upper thoracic region, i.e., above the level of the outflow of the greater part of the sympathetic fibers, abolishes physical heat regulation but leaves chemical regulation to a

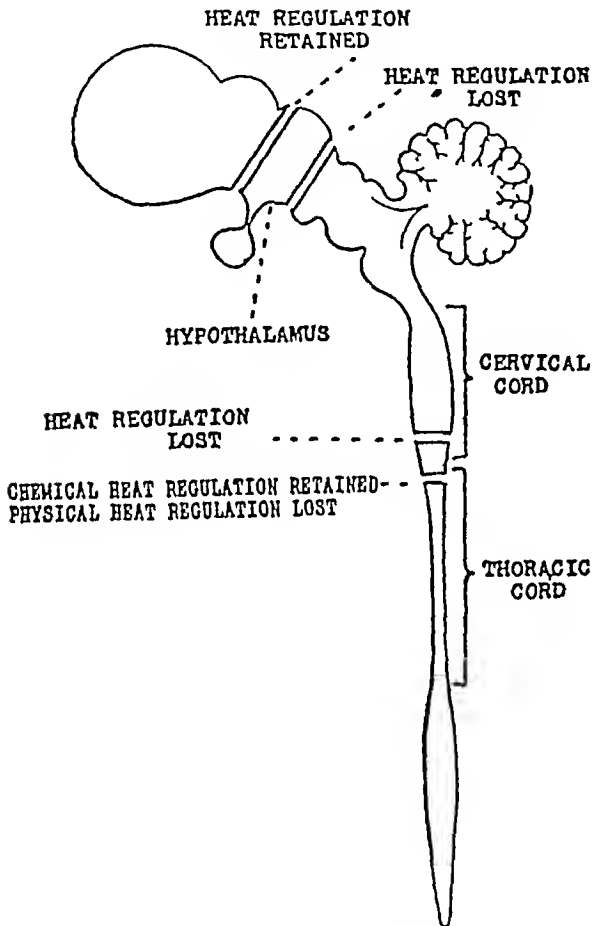


FIG 53 5 Diagram to illustrate the nervous control of the heat regulating mechanisms (Modified from Martin)

large extent intact, since the muscles of the fore part of the body remain in communication with the central nervous system. The effect of section through the brain stem indicates that the main center or centers controlling heat regulation must lie anterior to the superior colliculi. Removal of the cerebral cortex, thalamus or corpus striatum does not destroy the controlling mechanisms so long as the hypothalamus remains intact. Destruction of this region alone, however, was found by Keller and Hare to abolish the ability to main-

tain a normal body temperature upon exposure to cold. The heat-dissipating mechanisms remained intact and were, in fact, released from restraint by the hypothalamic lesion, as evidenced by pronounced panting, vasodilatation, and sweating of the toe pads (of cats).

The thermoregulatory centers have been located more precisely by Ranson and his associates. From the results of their experiments upon cats and monkeys they place the centers controlling heat loss (sweating and panting) in the preoptic and supraoptic regions between the anterior commissure and the optic chiasma. Heating this area causes sweating and panting or rapid breathing and a fall in body temperature. Destructive lesions in this location are followed by hyperthermia when the animal is exposed to a degree of heat that would cause little effect upon the body temperature of a normal animal. Sweating or panting does not occur though the temperature rises to over 106°F. Such lesions have little effect upon the animal's ability to resist cold. The centers controlling heat production and heat conservation, i.e., the mechanism whereby an animal is enabled to maintain a normal body temperature when exposed to cold, is situated, according to these investigators, in the caudal part of the lateral hypothalamus; it appears to be identical with the sympathetic center. The existence of a center for shivering (heat production) in the hypothalamus is suggested by the fact that under certain circumstances shivering in animals is accompanied by some of the manifestations of "sham rage" (Barcroft and Izquierdo). That the posterior part of the hypothalamus contains the main shivering center is indicated by the experimental results of Kellar and his associates and of Ranson and his colleagues who found that, in cats, shivering was abolished by a destructive lesion in this location. The efferent pathway for shivering is unknown, but an observation upon a patient in whom both spino-thalamic tracts had been sectioned and who did not shiver when the legs were immersed in cold water, suggests that the impulses descend the cord by these paths. The impulses are not transmitted by sympathetic nerves, since shivering occurs in sympathectomized parts. Shivering is reduced or abolished by certain drugs, especially calcium chloride and aminopyrine which most probably act upon the shivering center.

From observations of patients with intracranial lesions involving the base of the brain, it seems

most likely that in man the centers are situated as described by Ranson and his associates for animals. Lesions in the supraoptic region are sometimes associated with hyperthermia, hypothermia, on the other hand, may accompany lesions involving the posterior part of the hypothalamus.

The main heat-regulating centers are apparently influenced in two ways—reflexly from the skin and by the temperature of the blood flowing through them.

The posterior hypothalamic center apparently exerts its controlling influence upon temperature through the transmission of sympathetic impulses to the cutaneous vessels, sweat glands and pilomotor muscles, removal of the sympathetic renders an animal unduly susceptible to cold (ch 72).

THE ENDOCRINES IN THERMOREGULATION

The thyroid and adrenals also play their rôles, probably important ones, in the regulation of body temperature. The calorogenic effects of the secretions of these glands are well known (chapters 58 and 59). Cannon observed that exposure to cold caused an increase in the rate of the denervated heart. It has also been reported that the serum taken from an animal exposed to cold raises the metabolism of a second animal into which it is injected. If the first animal has been thyroidectomized, the effect upon the metabolism of the second is not observed. The adrenal secretion exerts a calorogenic effect which is immediate and of short duration. Its liberation follows short periods of exposure to cold. But owing to the delayed action of the thyroid hormone, and the persistence of its effects, it does not seem possible that the thyroid could play any part in increasing heat production unless the cold stimulus were continued over a long period. Rats exposed over a period of weeks to low temperatures (7.8 to 12.2°C) showed thyroid hyperplasia and a rise in metabolic rate of as much as 16 per cent which was not reached, however, until the lapse of from 2 to 4 weeks. Thyroidectomized rats, on the contrary, show little rise in metabolic rate under the same conditions. The experiments of Uotila point to the pituitary as being responsible (influenced through the hypothalamus) for the thyroid response to cold, for hypophysectomy causes thyroid atrophy as usual, though the animals are exposed to a low temperature for long periods. It may be

mentioned in this connection that the temperature tends to be subnormal in suprarenalectomized or thyroidectomized animals, in Addison's disease, and in cretinism.

DISTURBANCES OF HEAT REGULATION

Since the body temperature represents the balance struck between heat production and heat loss, an alteration in the value of one of these factors in relation to the other obviously will be followed by a change in body temperature.

Temporary rises in temperature may occur in health as a result of hot baths, which prevent heat loss through conduction, radiation and the evaporation of sweat, as well as by actually adding heat to the body, or during violent muscular exercise which increases heat production. Owing to the high efficiency of the chemical mechanism of control, a fall in the temperature of a normal person is much more difficult to produce than is a rise.

The heat regulating mechanisms are depressed by anesthetics, during sleep and hypnosis and by general bodily fatigue.

HEAT STROKE AND HEAT EXHAUSTION

In *heat stroke* the subject, previously well, suddenly becomes unconscious. Cessation of sweat secretion immediately precedes the loss of consciousness. The body temperature (rectal) is found to be around 110°F (Waterlow). There appears to be a complete breakdown of the heat regulating mechanisms. The hyperthermia may seriously damage the nervous centers and prove fatal. The blood chloride is reduced, but the reduction is due to dilution, for the plasma volume is considerably increased. The urine volume is increased, and chloride is excreted in normal amount.

Heat exhaustion develops more gradually than does heat stroke. The subject suffers from abdominal cramps and vomiting, and shows signs of circulatory collapse. When standing his pulse is rapid and his systolic blood pressure low. The plasma volume is reduced, sometimes by as much as 50 per cent. The diastolic pressure in the recumbent position tends to be higher than normal due to peripheral vasoconstriction, which is evidently a compensatory response to the reduced blood volume. The urine is of very small volume and is almost or quite free from sodium chloride, the blood chloride is greatly reduced due to the loss of salt in the sweat for some time preceding the onset of the symptoms. The amount of salt lost in the sweat may exceed 25 grams daily. According to Waterlow, heat exhaustion is seen also in another form, the chief manifestations of which are dizziness, palpitation, dyspnea, insomnia, a cutaneous rash and subnormal sweating.

Sunstroke is simply a form of heat stroke or heat exhaustion, but in addition to the reduction in heat loss as a result of the high atmospheric temperature there is an absorption of solar radiant energy. This may cause a local elevation of temperature above that of the body generally in regions such as the brain or cervical cord, which are unprotected from the heating effects of the sun's rays. Powerful sunshine itself, however, will not cause sunstroke provided heat dissipation is adequate to keep pace with heat production, i.e., when the air is dry and cool and strenuous exercise is not undertaken.

FEVER (PYREXIA)

Types of fever

(1) *Infectious fever*, e.g., sepsis, typhoid, pneumonia, etc.

(2) *Surgical fever* which arises after an extensive aseptic operation and is apparently due to toxic substances liberated by the injured tissues.

(3) *Neurogenic fever* from injuries to nervous centers, especially lesions in the neighborhood of the third ventricle, internal capsule, medulla or upper part of spinal cord.

(4) *Fever of dehydration* due to a reduction of blood-water (anhydremia, p. 24). This is particularly likely to occur in young children.

(5) *Fever produced by drugs and other chemical substances*.

Body temperatures as high as 113°F have occasionally been reported, but survival with a hyperpyrexia above 112°F is rare. In the great majority (over 95 per cent) of fevers, from whatever cause, the temperature does not exceed 106°F. Temperatures higher than 107°F become harmful or dangerous from the high temperature itself. The infrequency of temperatures above 106°F suggests that some safety thermostatic mechanism comes into operation around this level.

Intravenous injections of concentrated solutions of glucose or salt induce fever by causing anhydremia (p. 24). Drastic cathartics, by drawing water from the blood into the bowel, may cause fever in a similar manner. Caffeine and cocaine in large doses induce fever by increasing muscular tone (greater heat production) and by causing blood concentration (reduced heat loss, see also p. 731). Hemoglobin solutions when injected into the blood stream exert a pyretic action, the hemolysis resulting from the intravenous injection of distilled water acts similarly. The manner in which the fever is produced is unknown. Beta-tetrahydronaphthylamine injected subcutaneously raises the temperature by its action upon the central and peripheral sympathetic mechanisms, causing cutaneous

vasoconstriction and consequently a greater conservation of heat. It also, through its action upon the muscles, causes increased heat production. The adrenal medulla is also probably stimulated by this drug and the outpouring of adrenaline may be an added factor in the temperature rise. Ergotoxine causes a rise in temperature in some animals (cat), presumably through a direct action upon the heat centers. Dinitrophenol, a drug sometimes used in the treatment of obesity (p. 710), and injections of foreign protein also raise the body temperature. Dinitrophenol acts by stimulating oxidative processes in the tissues. Adrenaline and thyroxine in large doses may also, through their stimulating effect upon the metabolism, cause a rise in temperature.

Infectious fevers

At the onset of an acute infectious fever the heat balance is upset by a reduction in heat loss as a result of vasoconstriction and a reduction in blood volume, combined with an increase in heat production. That is to say, those mechanisms which in health prevent a fall in temperature when the body is exposed to cold, are called into play by a stimulus within the body itself, namely, the toxin of the infecting organism.

The reduced heat loss by radiation and convection in the early stages of the fever is evident in the cold, pale or slightly cyanosed skin. At this time, though the body temperature may be higher than at any subsequent stage, the patient often experiences sensations of extreme cold (chills) with shivering, chattering of the teeth and "goose flesh." In health, the comfortable feeling of warmth depends not upon the temperature of the deeper structures but upon the stimulation of the cutaneous sense organs (corpuscles of Ruffini) by the warm blood coursing through the superficial vessels. The chills are due to the spasm of these vessels and the exclusion from them of the warm blood of deeper regions. The fall in skin temperature acts as a stimulus which calls into play the mechanism of chemical regulation. Shivering, which consists of fine contractions of the muscles, occurs, muscle tone increases, and the smooth muscle of the skin contracts. The increased heat production thereby induced is an additional factor in the elevation of the body temperature. Later, when the body temperature reaches a certain height, a heat response is evoked from the centers, the vessels are released from spasm, the blood flow through the skin increases, the body surface becomes flushed, and the patient feels intensely hot. The balance between

heat loss and heat production is again restored but set at a higher level than in health. The body's "thermostat" is turned up a point or so.

DuBois, in an experiment upon a normal man and a malarial fever patient, demonstrated the reduced heat elimination which occurs during the chill. The normal subject imitated as closely as possible for a period of 34 minutes the shivering of the patient and thereby increased his heat production by nearly 200 per cent. Most of the extra heat was eliminated as it was produced, the body temperature showing only a slight rise. In the malarial patient, on the contrary, in whom the heat production during the chill was increased to about the same degree, all the extra heat was retained.⁸ The heat retention caused a rise of 2°C in body temperature. After the chill the heat elimination rose and the temperature fell.

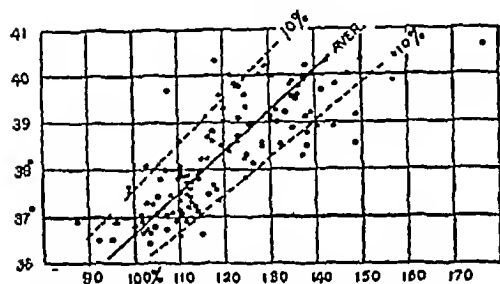


FIG 53.6 Relationship between basal metabolism and body temperature in fever. Results in six different fevers grouped in one chart. The continued line shows the average metabolism, the interrupted lines the metabolism 10 per cent above and 10 per cent below the average respectively (After DuBois).

The *continued* fever, which usually follows the initial chill of an infectious disease, or which develops in other instances without this preliminary, is due essentially to the raised threshold for heat loss. There is, of course, increased heat production but this is mainly secondary—the velocity of the oxidative processes of the tissues being increased by the rise in temperature induced by the diminished heat elimination.⁹ That the latter,

⁸ The water equivalent of the human body is 0.83 times its weight in kilograms (W) per °C rise in temperature (t°C), the retained heat may therefore be calculated from the formula $t^{\circ} \times W \times 0.83$.

⁹ The increase in metabolism with rise in temperature follows van't Hoff's law which states that the velocity of chemical reactions increases from 2 to 3 times for a temperature rise of 10°C. That is, the temperature coefficient is between 2 and 3. The basal metabolism actually increases about 13 per cent for each degree Centigrade rise in body temperature or 7 per cent for each degree Fahrenheit (fig. 53.6).

rather than increased heat production, is the principal factor in the elevated temperature is indicated by the fact that in fever a temperature of 104°F (40°C) is accompanied by an increased heat production of only about 35 per cent, whereas in health the heat production must be increased several fold (as in muscular exercise) in order to raise the body temperature even transiently to this extent. Also, the metabolism may be raised 40 per cent or so in an animal by protein feeding (see specific dynamic action, p. 640) without the occurrence of a temperature rise. Furthermore, the increased metabolism in ordinary fevers occurs simultaneously with the temperature rise. If the latter were the result rather than the cause of the former, it should be possible to demonstrate by indirect calorimetry a period of increased metabolism *preceding* the rise in temperature, this, however, does not occur.

With the termination of the fever, sweating (or at least moistening and cooling of the skin) occurs, the heat balance is restored to its normal level, the heat which had been retained is eliminated either gradually or with comparative abruptness (crisis) and the temperature falls.

In infectious fevers the bacterial toxins act either directly or indirectly upon the heat center or centers. In decerebrate animals, or in those in which the cord has been divided in the lower cervical or upper thoracic regions, the injection of infective agents does not cause fever nor is a rise in temperature so produced in animals possessing no heat regulating mechanisms. Barbour suggests that the first effect of the toxin is not directly upon the heat regulating centers but upon the peripheral tissues. In these, it is conceived, chemical changes are initiated which result in the withdrawal of water from the blood. The reduced volume of circulating fluid causes blood to be drained from the skin vessels and thereby calls forth a reflex cold response from the heat regulating mechanisms. That is, the nervous centers respond by causing vascular spasm, *shivering* and a further increase in blood concentration. The mechanism involved at the onset of fever is therefore comparable to that which produces a rise in temperature of a healthy man after a cold plunge. The cold water causes, through vasoconstriction, diminished heat dissipation and through shivering, increased heat production. The body temperature may rise for a few minutes. After the first effect of the plunge the vessels are released from spasm, the stored heat is eliminated, the surface of the body glows and a feeling of comfortable warmth is experienced. This constitutes the "healthy reaction" of a cold shower and may be compared to the period following the fever chill, except

that in the latter case the heat balance is set at an abnormally high level. The factors causing the continuance of the fever, on the other hand, are similar to those which cause a rise in temperature during the immersion of the body in a hot bath. Heat dissipation is greatly reduced and heat production is increased purely as a result of the elevation in temperature of the tissue cells.

Special metabolism in fever

Water and salt At the onset of a high fever the blood volume, as already mentioned, is reduced, the reduction being due to the loss of water (anhydremia). Later the volume of the blood tends to increase (hydremlia) as a result of the shift of water from the tissues to the vessels. During the course of the fever the urine volume is markedly reduced, but is increased above the normal when the temperature falls. The vaporization of water from the skin and lungs is increased in fever, owing to the high temperature itself and to the more rapid breathing.

There occurs a retention of chloride which apparently is deposited in the tissues, the chloride concentration of the blood being normal or below normal. The urinary chlorides in most fevers are greatly reduced. At the termination of the fever the retained water and chloride are eliminated by diuresis and sweating. Chloride retention is especially pronounced in pneumonia.

Protein The excretion of nitrogen in the urine is greatly increased in most infectious fevers. This is furnished by body protein, the protein minimum, i.e., the "wear and tear" quota of protein metabolism (p. 638), being much higher than normal. In very severe infections from 300 to 400 grams of body protein may be destroyed daily. It has been found impossible to maintain the fever patient in nitrogen equilibrium (p. 638) by giving liberal allowances of protein combined with supplies of carbohydrate which under ordinary circumstances would be considered quite adequate for energy purposes. It has, therefore, been held that the toxins of the disease were responsible for the protein destruction, the so-called "*toxic destruction*" of protein. It has been shown, however, by Shaffer and Coleman that if a diet be given possessing a caloric value 50 to 110 per cent in excess of the patient's requirements, as actually determined by calorimetry, and containing a liberal supply of protein (160 to 200 grams), nitrogen equilibrium can be established. The high protein catabolism which has been observed in

fever patients on the usual diet is therefore thought to have been due in large measure to the fact that the caloric intake was far below the requirements, which owing to the higher metabolism in fever are considerably greater than has been supposed. In other words, a fever patient upon a diet which has been considered adequate in the past is actually in a half-starved state, and is, in consequence, forced to consume his own tissues.

Nevertheless, even on a high caloric diet composed of carbohydrate and a small quantity of fat, the nitrogen excretion still remains well above that of a normal person. Moreover, a protein allowance equivalent to that of a healthy man, together with carbohydrate somewhat more than sufficient to cover the calculated caloric requirements of the febrile state, will not maintain nitrogen equilibrium. As just mentioned, the caloric allowance must in some instances be double the heat production of the patient. It therefore appears that the toxic process itself must be responsible in part for the increase in protein metabolism. Creatinine, uric acid, purine bases and phosphates also appear in the urine in increased amounts—further evidence of a destruction of body protein. The manner in which the "toxic" effect is produced is not known. It does not appear to be due merely to the high temperature, for, raising the temperature of a normal person to 104°F by immersion in a hot bath does not increase the nitrogen excretion significantly. In fevers with much destruction of body protein the specific dynamic action of food is absent.

It must be apparent from the foregoing remarks that in order to reduce the waste of body tissue in fever a liberal diet should be given, provided such is not contraindicated by some special feature of the disease. The old adage "feed a fever" holds true. Since the very high protein diet required for the establishment of nitrogen equilibrium in the fever patient has its disadvantages, one must usually be satisfied with reducing the waste of body protein rather than aiming to abolish it. Special attention, therefore, is directed toward furnishing an abundance of protein-sparing food (p. 639), namely carbohydrate, and thus avoiding excessive quantities of protein. The more abundant diet causes a negligible increase in heat production and no elevation of the temperature.

Fats and carbohydrates The metabolism of body fat or carbohydrate shows no definite abnormality in fever. In patients upon a low food intake

body fat and glycogen are utilized as fuel. Acidosis results from the excess combustion of fat only if the available carbohydrate is inadequate in amount.

THE ACTION OF ANTIPYRETIC (FEVER REDUCING) DRUGS A list of chemical substances which are capable of inducing a rise in body temperature has been given on page 739. Other drugs, e.g., antipyrine, aspirin, salicylates, quinine, etc., though they exert little effect upon the normal temperature, lower the temperature in fever by increasing heat elimination. According to Barbour they bring about this effect through drawing water from the tissues into the vessels and thus increasing the volume of fluid in the body's heat-radiating system. They appear to exert little effect upon heat production. The effect upon the blood volume is possibly brought about indirectly. All these substances raise the blood sugar, the greater sugar concentration may then through osmotic forces attract water into the vascular system.

Certain other drugs, e.g., morphine, general anesthetics and alcohol, tend to depress the *normal* body temperature chiefly through blood dilution and dilatation of superficial vessels. In the case of morphine and general anesthetics, a direct depressant action upon the heat centers is also indicated.

THE VALUE OF FEVER. Fever is frequently the herald of serious disease, nevertheless, unless of high degree and on this account endangering the functions of vital tissues, it should not be looked upon as a reaction detrimental in itself. On the contrary, there is every indication that its occurrence is an important aid to the body in its combat with the disease. The rôle which fever plays in the defensive process is, however, unknown. It has been suggested that the formation of antibodies can be elaborated only at higher temperatures. It is well known, for instance, that in infections which overwhelm the individual the temperature reaction is depressed. In support of the belief that moderate fever is not injurious in itself but is actually beneficial the following observations upon animals may be cited. (a) The body temperature of rabbits has been maintained by the application of external heat at a level of over 105°F for weeks at a time with

out ill effects. (b) In animals infected with certain microorganisms the disease runs a milder course when the temperature is raised (to 40°C) artificially. High temperature is thought to be favorable for the elaboration of antibodies. (c) It has been reported that with moderate overheating the formation of various antibodies is increased, but at higher temperatures the process, apparently, is depressed. The immunity of fowl to the ordinary pyogenic infections is ascribed to their higher body temperature which is inimical to the growth of pus-forming bacteria. (d) Fevers induced artificially by means of foreign proteins or injections of malarial blood are used as therapeutic agents in arthritis and in chronic nervous disease due to the pathogen of syphilis (*Treponema pallidum*). Short wave diathermy has been employed with success in the treatment of certain infections—the high temperature produced in the tissue exerting a lethal action upon the microorganisms. (e) In the past it has been common practice to reduce fever (above 104°F) by cold bathing, but it was found that the patients did not do well and the practice, except in extremely high fevers, has been abandoned.

REFRIGERATION—CRYMOTHRAPY In chronic incurable disease, e.g., carcinoma, refrigeration of the patient is sometimes resorted to for the relief of severe, intractable pain. Ice is packed around the patient or some other means is used to reduce the temperature of the body to between 80° or 90°F. This temperature is maintained for hours or days. The patient becomes unconscious, passing into a state of "suspended animation", which in many respects is comparable to that of an animal during hibernation. The heart rate and respirations are slowed and the radial pulse may be imperceptible. The blood pressure falls below the level at which it can be measured, urine production is greatly diminished or suppressed and gastrointestinal activity is minimal. The blood volume is reduced as a result of the loss of water (hemoconcentration or anhydremia). The metabolism, in accordance with van't Hoff's law, is reduced by from 20 to 50 per cent. The respiratory quotient is probably not greatly lowered, but in hibernating animals, in which the metabolic processes are altered qualitatively as well as quantitatively, it is around 0.60.

CHAPTER 54

THE VITAMINS

INTRODUCTION NOMENCLATURE

In 1911, Casimir Funk, as a result of his investigations into the cause of beri-beri, obtained a crystalline substance from rice polishings which was capable of preventing or curing this nervous disease. He named the substance *vitamine* in view of its quite evident importance to life, and believing, though erroneously, that it was an amine. Subsequently a number of similar substances playing an essential rôle in nutrition were discovered for which the general term *vitamin* was agreed upon, the individual vitamins being designated by letters of the alphabet. In the state of knowledge at the time such a method of designation had the advantage of being non-committal as to the chemical nature and other properties of this group of substances. The following is a classification of the known vitamins and their chief physiological effects.

Vitamin A (antixerophthalmic)

thiamin¹ (B₁)
riboflavin (B₂)
nicotinic acid
pyridoxin (B₆)
folic acid

Vitamin B complex

antipernicious anemia factor, B₁₂
pantothenic acid
inositol
biotin
para-aminobenzoic acid
choline

Vitamin C (antiscorbutic)

Vitamin D (antirachitic)

Vitamin E (antisterility)

Vitamin K (antihemorrhagic)

Vitamins A, D, E and K are fat-soluble. The others, namely, those of the B complex, and vitamin C are water-soluble. Riboflavin is also called vitamin B₂ and sometimes vitamin G (see p 752).

MODIFYING FACTORS IN THE ACTIONS OF VITAMINS

Antivitamins, toxamins. Eijkman, many years ago, as a result of his pioneer experiments on polyneuritis in fowl (p 749), postulated that the nervous disease was caused by a toxic principle in polished rice which was antidoted by a second

factor present in the coverings (polishings) of the rice kernel removed in milling. But after Hopkins showed that nutritional defects resulted from deficiencies of essential dietary elements, the notion of a toxic substance in food was forgotten until Mellanby, some years later, demonstrated the anticalcifying and rachitogenic properties of cereals. He suggested that these effects were due to the presence of a toxic principle which antagonized the action of vitamin D. He named it *toxamin* (see phytic acid p 767 and p 768). A toxic principle in maize, also first postulated by Mellanby, has been shown to antagonize the action of nicotinic acid. Several other substances antagonistic to the action of vitamins have since been recognized. Thus, carbohydrate, in a sense, is an antivitamin for thiamin (p 750), pigs fed upon yeast, which is rich in high quality protein and vitamins of the B complex, become rachitic unless the diet is supplemented with vitamin D. A rachitic factor in young green cereals used in New Zealand as sheep fodder has been demonstrated, its effect is prevented by vitamin D, but not by exposure to sunlight. The actions of vitamins A and E are antagonized by rancid fats (p 746), and vitamin K by dicumarol (p 118). Other examples of antivitamin action are the inactivation of biotin by avidin (p 756), of folic acid by aminopterin (ch 9) and certain others of its analogues, a substance in ergot of rye which, as shown by Mellanby, antagonizes the action of vitamin A, and an enzyme factor (thiaminase) in fresh, raw fish which destroys thiamin, ranch foxes, upon a diet containing 20 per cent of raw fish, develop a paralysis which is cured by added thiamin. Several chemical analogues of vitamins, owing to their structural resemblance, are antagonistic to certain vitamins, competing with them for a position in some vital metabolic process. The outstanding example of this action is the antagonism between sulfanilamide and para-aminobenzoic acid (PABA) a factor of the B complex. The drug owes its bacteriostatic effect to its affinity for the same component of the bacterial cell with which PABA normally reacts (p 756). Most of the analogues of vitamins shown to be antagonistic are artificial products and are therefore of little practical importance. Examples

¹ Also spelled thiamine

of such antagonisms are those between *pyridoxamine* and thiamin, between *pantoylaurine* and pantothenic acid, between riboflavin and *6,7-dichloro-9-ribityl-isoxanthine* and between pyridoxine and a structurally similar compound containing tellurium in place of nitrogen in the pyridine ring. The action of vitamin C (ascorbic acid) is antagonized by *glucoascorbic acid*. When this compound is administered to guinea pigs, a scurvy like condition develops but can be completely prevented if sufficient amounts of ascorbic acid are given at the same time.

Hypervitaminosis—vitamin imbalance. Certain vitamins when given in excessive dosage give rise to toxic manifestations. These, in some cases at least, are thought to be due to a disturbance in the vitamin balance. Overdosage with vitamin A, for example, may cause a hemorrhagic state suggestive of scurvy, irregular thickening of the cortex of bone, sometimes with hyperostoses, and general symptoms, loss of appetite and muscular weakness, rarefaction and fragility of the bones have also been observed. Massive doses of vitamin D have long been recognized to be followed, especially in children by intensely toxic effects (p. 857). Thiamin in excessive amounts may give rise to symptoms and signs resembling those of thyrotoxicosis, e.g., nervousness, tremor, tachycardia and sweating. Or anaphylactoid effects may be produced, e.g., asthmatic attacks, dyspnea, urticaria, nausea etc. Death from a state closely resembling anaphylactic shock has also been reported as a result of overdosage with this vitamin.

Of the other vitamins, naturally occurring vitamins K₁ and K₂ are non-toxic even in massive doses, but the synthetic naphthoquinone and naphthohydroquinone compounds cause porphyria and vomiting, the first of these in large doses *prolongs* the coagulation time. Riboflavin, pantothenic acid, inositol, nicotinic amide and biotin, and vitamins C and E, appear to be entirely free from toxicity. Nicotinic acid and pyridoxine may produce mildly toxic effects, but only in extremely high dosage.

VITAMIN A (ANTIXEROPHTHALMIC)

This was discovered as a result of the investigations of Hopkins in England and of Osborne and Mendel, and McCollum and Davis in America.

Sources

The chief sources of vitamin A are mammalian and fish liver, egg yolk, butter, cream and a num-

ber of vegetable foods. Cod liver oil has a very high vitamin A content, and halibut oil a much higher one. Vegetable oils with the exception of corn oil and red palm oil contain little or none. Lard and beef fat are, as a rule, poor sources. Their vitamin A content varies considerably, however, with the animal's diet. Cereals, with the exception of maize, are relatively poor in vitamin A. In plant tissues there is a definite relationship between their green or yellow coloring and their vitamin A activity (see below). Thus, the sweet (yellow) potato is a good source, whereas the ordinary potato possesses very little or none, the outer green leaves of lettuce contain some 30 times more of the vitamin than do the inner white leaves. Carrots, yellow maize, escarole, spinach, cress string beans, green peas, pumpkin, bananas and cantaloupe are rich in the vitamin, whereas in white corn, celery, cauliflower, white turnips, cabbage, radishes and other colorless vegetables it is present in small amounts or entirely lacking.

The daily *vitamin A requirement* for adult man is from 5000 to 6000 international units, and from 1500 (under 1 year) to 5000 (at 15 years).

Chemical properties and history in the body

Vitamin A is soluble in fats and fat solvents. It is present in the unsaponifiable fraction of the fat. It is resistant to heat in the absence of air but is readily destroyed by oxidation at all temperatures. Though colorless itself it gives a blue color with antimony trichloride in the presence of oxygen. In oily solution, vitamin A, or its esters in alcoholic solution, show a green fluorescence when exposed to ultraviolet light. This property has been employed to demonstrate microscopically its distribution in the body tissues, e.g., Kupffer cells, retina, adrenal cortex, testes and corpus luteum. It is also seen in actively secreting mammary glands and in tumor tissue. The fluorescence gradually fades, presumably as a result of the destruction of the vitamin by the ultraviolet rays.

Vitamin A is an unsaturated alcohol with the empirical formula $C_{20}H_{30}O$ and is derived from the orange yellow pigment *beta-carotene*— $C_{40}H_{56}$ —one molecule of the latter being split into two molecules of the vitamin, thus, $C_{40}H_{56} + 2H_2O \rightarrow 2C_{20}H_{30}O$. Vitamin A has been isolated in crystalline form and was synthesized by Milas in 1946, its formula is shown on page 745. Euler, in 1928, showed that pure carotene from carrots was car-

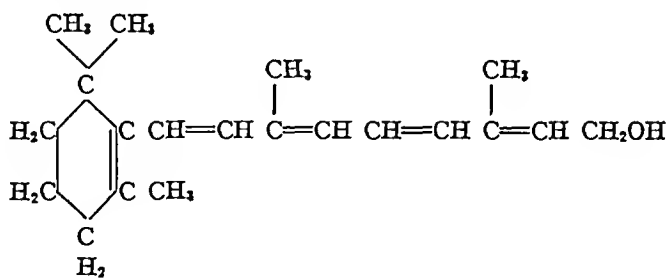
² Carotinoid pigments of vegetable food are also largely responsible for the fat pigment of animals.

pable of replacing vitamin A in the diet. Moore later found that feeding rats with pure carotene caused a great increase in the vitamin A contents of their livers. It is now established that carotenes, of which there are three forms, *alpha*, *beta* and *gamma*, and possibly other plant pigments (e.g., cryptoxanthin) are the precursors of vitamin A. These pigments are therefore called *provitamins A*.³ Beta-carotene is of chief importance, the alpha form gives rise to one molecule only of vitamin A. The vitamin is stored mainly (95 per cent) in the liver, but also in the adrenal cortex, retina, corpus luteum and body fat. The polar bear stores enormous amounts—26,000 international units (per gram)—in its liver. From the observations of Moore it has been presumed that the carotene-vitamin conversion also takes place mainly in the liver (Kupffer cells), but the evidence for this is not entirely satisfactory. Olcott and McCann obtained evidence (from ultra violet absorption spectra) that vitamin A was formed from carotene when incubated with liver slices. The conversion was thought to be catalyzed by an enzyme, *carotenase*. This observation has been both confirmed and denied. Drummond and McWalter, though they found an increase in the vitamin A content of the livers of rabbits eight days after the intravenous administration of caro-

of vitamin A deficiency. The possibility is suggested that the carotene-vitamin A conversion takes place in the intestinal wall. The carotene-vitamin A transformation is a relatively slow process and not complete, probably not more than from 30 to 50 per cent of ingested carotene is converted into the vitamin.

Vitamin A does not exist as such in plants but only in the form of the provitamin.⁴ Algae, diatoms and other marine plants synthesize the provitamin, and serve as food for small marine forms referred to in general as zooplankton, i.e., copepods, molluscs, etc. The zooplankton serve as food for various species of small fish (herring, whiting, young cod, etc.) which thus receive the necessary supplies of provitamin. The smaller fish upon which the larger fish such as cod and halibut feed provide these latter with vitamin A. Herbivorous animals obtain the provitamin from fodder (alfalfa is a particularly rich source). Dairy cattle convert it in part into the colorless vitamin A.⁵ This, as well as unchanged carotene, is secreted in the milk. Man may therefore acquire a supply of this vitamin by consuming either plant foods rich in the provitamin or animal tissues which contain it preformed.

The concentration of vitamin A in the liver of the new born infant is low, for the placenta transmits little from the mother. The concentration in-



Vitamin A₁

tene, did not observe any transformation of carotene to vitamin A *in vitro*. More recently Sexton and Deuel and their associates have shown that, whereas carotene given orally was followed by an increase of vitamin A in the liver, when given parenterally, it accumulated in the liver and no increase in the hepatic stores of vitamin A was observed. Carotene administered parenterally was also relatively ineffective in correcting the effects

creases gradually during the first few weeks of life, the vitamin being supplied to the off-spring in the milk. The quantities of vitamin A and carotene in human milk are relatively high, and are highest in colostrum and early milk. Bile, though necessary for the absorption of carotene, is not required for the absorption of vitamin A. Absorption of both

⁴ Since the actual and only important effect of carotene after ingestion is that of producing vitamin A, it still remains the custom to speak of the vitamin A value of a food rather than of its carotene content.

⁵ The proportion of carotene converted, and so the color of the milk, varies in different breeds of cattle. It is evident that a pale-colored milk may have a vitamin A value as high as or higher than one more richly colored (due to carotene).

³ The provitamins are synthesized in the leaves and stalks of plants under the influence of sunlight. The high content of carotene in root structures, e.g., carrots, swedes, etc., is difficult to explain unless it is the result of transference as some water soluble precursor from the leaves to the roots.

the vitamin and provitamin on the other hand, is favored by the presence of fat or phospholipids, e.g., lecithin, in the intestine, whereas petroleum has a definitely depressing effect, especially upon the absorption of carotene. The vitamin is found in the liver and intestinal tract, chiefly in the form of esters which before their absorption from the latter situation are hydrolyzed by the lipases of the intestinal and pancreatic juices. During absorption the vitamin is found in the intestinal mucosa in the form of the alcohol.

In experiments upon rats Moore discovered an interesting synergic relationship of vitamin E with vitamin A. The administration of vitamin E increases the storage of vitamin A in the liver, prolongs the time required to exhaust the reserves and postpones the central abnormalities due to vitamin A deficiency. Vitamin E probably exerts its effect by acting as an antioxidant agent in the gastro-intestinal tract (p. 769). Rancid fats and unsaturated fatty acid esters (molecules isolated) have the opposite effect, tending to inactivate vitamin A through oxidation.

Vitamin A has been given the following structural formula by Karrer:

Below is a comparison of the properties of carotene and vitamin A.

Carotene	Vitamin A
Synthesized by plants	Stored by animals
Reddish-yellow in color	Almost colorless
Absorption bands at 325 m μ absent	Absorption bands at 325 m μ present
Greenish-blue color reaction with antimony trichloride giving absorption bands maximal at 590 m μ	Vivid blue color—antimony trichloride, absorption band maximal between 615 and 620 m μ

The manner in which vitamin A exerts its action is not known precisely, but it probably plays a rôle in the tissues as part of an oxidation-reduction enzyme system.

A second vitamin A, called vitamin A₂, is present in some fish-liver oils, but, except in those from fresh-water fish its concentration is very much lower than the vitamin derived from β -carotene. This latter form is now referred to as vitamin A₁. Vitamin A₂ is believed to differ chemically from A₁ in possessing an additional CH₃ group and an extra conjugate double bond. In antimony trichloride it shows an absorption band maximal at 696 m μ . It is the predominant form found in the

livers of fresh-water fish but has not been obtained from the livers of mammals.

The relation of vitamin A₁ to rhodopsin and of A₂ to porphyropsin is outlined in chapter 74.

The importance of vitamin A in rhodopsin and rhodopsin, rhodopsin

The following effects result from vitamin A deficiency, which may be due not only to a dietary lack but also to failure in absorption, as in obstructive jaundice and sprue, or to severe liver disease (e.g., cirrhosis) in which the carotene conversion process is impaired.



FIG. 541. Illustrates a baby which had suffered from an acute xerophthalmia of dietary origin, and was cured by administration of fat-soluble A as butter and cod-liver oil. The disease had however progressed so far that the eye of the left eye was destroyed and the right eye damaged. (After Black, from McCollum and Simmons *The Newer Knowledge of Nutrition*.)

(a) *Dermatitis of the skin*. In man one of the earliest manifestations of vitamin A deficiency is dryness of the skin followed by a papular eruption due to changes in the hair follicles, the sebaceous glands and sweat glands atrophy. Hyperkeratosis and the formation of keratotic plugs in the hair follicles are seen in man and in rats.

(b) *Xerophthalmia and inflammatory eye conditions*. In xerophthalmia the primary change appears to be in the lacrimal glands whose secretion is suppressed. The corneal surface becomes dry, and, having lost the protective and lubricating effect of the tears, becomes invaded by microorganisms (fig. 541). Inflammation and thickening of the conjunctivae with a purulent discharge,

and softening (keratomalacia) leading to ulceration of the cornea result. Osborne and Mendel observed this condition in 80 per cent of rats placed upon a vitamin A deficient diet. It occurs in the human subject when the diet is lacking in this vitamin and there is no doubt that it is a specific manifestation of such deficiency, since it cannot be produced in animals by the lack of any other vitamin.

(c) *Cornification of epithelial surfaces* The epithelial linings of the respiratory, alimentary and urinary tracts and the ducts of various glands, tend to become converted to the stratified squamous type with consequent drying up of their secretions. Evans and Bishop observed that the cornified vaginal cells characteristic of the estrus period of the rat (p. 876) appeared in animals upon a diet deficient in vitamin A even though the ovaries had previously been removed. The changes in the lachrymal glands, hair follicles and cutaneous glands mentioned above, as well as several other manifestations are special examples of the general tendency toward cornification of epithelial tissues when the diet is deficient in vitamin A.

(d) *Night blindness (nyctalopia)* This is the failure of vision in dim light which occurs in man and animals as a result of vitamin A deficiency. It is not uncommon in the tropics, and an interesting example of custom anticipating science is the fact that the native treatment for the condition was a poultice of liver to the eyes and the addition of liver to the diet. Night blindness is also seen in Labrador and certain parts of Newfoundland where the diet in winter is deficient. The condition is due to the failure in regeneration of the visual purple (ch. 74) after the eyes have been exposed to bright light. In health, regeneration of the pigment occurs after the eyes have been a few minutes in the dark, whereas in vitamin A deficient animals (rats) the regeneration occurs very slowly or not at all.

It should be emphasized that vitamin A deficiency is not the only cause of night blindness, and even when it is the cause the visual defect cannot always be attributed to failure in the regeneration of visual purple, for degenerative changes in the visual receptors or of the neural elements of the retina may be a relatively early effect of vitamin A deficiency.

A method has been devised for detecting mild grades of vitamin A deficiency based upon the measurement of the rate of dark adaptation. The eyes are first

accustomed to complete darkness for a period of 10 minutes and then exposed to a bright light seen through the eyepiece of a specially designed photometer (or adaptometer) for 3 minutes. The light is then switched off and the rate of dark adaptation determined for a period 10 minutes. The clinical value of this test is in some doubt. The determination of the concentration of the vitamin in the plasma is a more sensitive and reliable test. Some believe that impairment of the faculty of dark adaptation, due to mild grades of vitamin A deficiency, may be a contributory factor in many motor accidents occurring at night.

(e) *Maldevelopment of bone* Vitamin A is an essential dietary constituent for the normal and orderly development of bony structures. In Mellanby's experiments on deficiency of this vitamin in puppies, gross abnormalities in bone growth were observed. The cancellous tissue is greatly thickened and the entire bone misshapen. The bones in relation to the central nervous system, the cranium (sphenoid, basi-occipital, and labyrinth) and vertebral column were studied especially because of the very serious effects which the osseous deformities exert on the neighboring nervous structures. But all the bones of the body were found to be affected more or less by vitamin A deficiency. The vitamin exerts a specific action upon the osteoblasts and osteoclasts of the growing bone. When it is lacking the normal balance between the activities of these two types of bone cell is disturbed. The site of their specific actions upon the bone—whether to provide thickening here or thinning there, as well as the degree of activity exhibited—appears to be uncontrolled. As a consequence, resorption and deposition of osseous tissue, which when properly directed leads to the molding of the bone into its characteristic shape, does not proceed in the normal fashion. In the case of the cranial and vertebral bones, growth and sculpturing of the enclosing bony tissue is not nicely co-ordinated with the development of nervous structures. Compression, distortion of nervous tissue, and bending or pinching of nerves in their passage through the foramina of the bones are a result.

(f) *Degenerative changes in the nervous system* Mellanby reported that there occurred in dogs upon a diet rich in cereals and deficient in vitamin A a condition resembling subacute combined degeneration of the cord in man (ch. 66). Wolbach and Bessey conclude, however, from their experiments upon growing rats that the effect of vitamin A deficiency upon the nervous system is not

direct, but is of mechanical origin due to the retardation of the growth of the bones of the vertebral column

There is some evidence that vitamin A deficiency in man in certain instances leads to optic neuritis. Degenerative changes in the layers of the retina including the outer portions of the rods have been described as resulting from vitamin A deficiency and it is probable that the changes in the rods are responsible, in many instances, for night blindness in vitamin A deficiency. Degeneration of other cranial nerves, especially of sensory fibers, were also observed in Mellanby's animals. The degenerative changes are indirect effects of malformation of neighboring bony structures.

Cervical Segment I

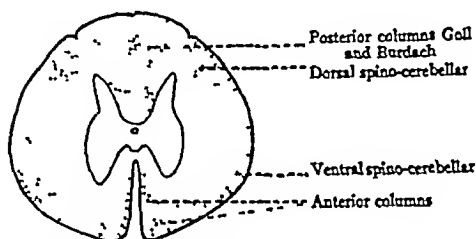


FIG 54.2 Degeneration in spinal cord of dog fed on a vitamin A-deficient diet. (After E. Mellanby, *Nutrition and Disease*)

(g) An antagonism between vitamin A and the thyroid hormone has been demonstrated. The loss in weight of rats caused by daily injections of thyroxine can be prevented by the administration of carotene. It has also been shown that the stores of vitamin A in the livers of guinea pigs are reduced by injections of thyroxine and the metamorphosis of tadpoles treated with thyroxine can be delayed by treatment with vitamin A. It appears that the thyroxine effect is due to its causing rapid destruction of the vitamin and not because it prevents the conversion of the provitamin. Indeed in *hypothyroidism* the conversion of carotene to the vitamin is defective, the milk of goats, which ordinarily is pure white, becomes yellow after thyroidectomy, due to the excretion of unchanged carotene.

(h) *Renal function* Pathological changes in the kidneys have been reported by several workers as resulting from vitamin A deficiency in animals as well as in man. Vacuolization and calcification of the cells of the convoluted tubules, cloudy swelling of the cells of the collecting tubules and hyperplasia and cornification of the epithelium of the renal pelvis have been described. The experiments of Herrin and Nicholls indicate that vitamin A plays an important rôle in

renal physiology. They demonstrated a reduction up to 40 per cent in the urea clearances of dogs upon vitamin A deficient diets. The clearances were restored to normal by the administration of adequate amounts of the vitamin. On the other hand, the urea clearances were raised 42 to 100 per cent above the normal in dogs receiving large doses of vitamin A.

(i) The relationship of vitamin A to urinary lithiasis is discussed on p. 870.

For a time it was thought that vitamin A had a specific anti-infective action. The administration of this vitamin in relatively large dosage was stated by some clinical investigators to lower the incidence of certain infectious diseases, as well as to reduce the morbidity and mortality of such diseases when given during their course. These results have not been confirmed. Vitamin A does not appear to have any specific effect in combating infectious conditions.

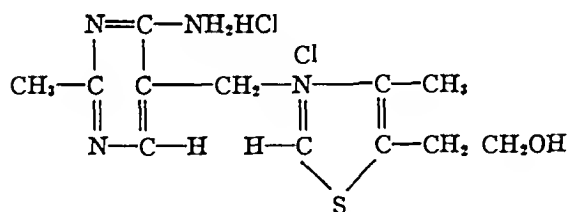
THE VITAMIN B COMPLEX

The first vitamin to be discovered was called the antineuritic vitamin since its absence from the diet was shown to be the cause of beriberi (see below). When the practice was adopted of referring to the vitamins by letter it was designated vitamin B. Later work showed it to be a combination of several vitamins, one is antineuritic and now known as vitamin B₁ or *thiamin*, the other components of the B complex are, *riboflavin*, *nicotinic acid* or *niacin*, *pyridoxin*, *pantothenic acid*, *inositol*, *biotin*, *para-aminobenzoic acid* and *choline*.

The grouping together of these water-soluble vitamins under the heading B complex is not artificial and simply a convenience, but based upon a common wide distribution and close association in vegetable and animal tissues, and upon their intimate functional interrelationships. They are of importance for the growth or metabolism of various forms of life from bacteria, protozoa and yeasts to mammals. They presumably form an essential part of the enzyme systems underlying living processes. Thiamin, riboflavin and nicotinic amide are present as prosthetic groups in certain tissue enzymes (ch. 32). Ten vitamin B factors have been obtained in pure form and a number have been synthesized in the laboratory, several are synthesized by the intestinal flora, e.g., thiamin, riboflavin, nicotinic amide, folic acid and biotin.

THIAMIN, VITAMIN B₁ (ANTINEURITIC)
OR ANEURIN

Chemical properties of B₁ This factor is soluble in water and in 70 per cent alcohol or acetone. It is a white, crystalline substance with a yeast-like odor and a characteristic nutty flavor. In acid media it is resistant to heat, remaining active for several hours after heating, with free access of air, to 100°C. It is, however, less resistant to heat than B₂, being destroyed at high temperatures (autoclaving), and is therefore spoken of as the *heat-labile* factor of the B complex, the B₂ factor being referred to as the *heat-stable* factor. B₁ is resistant to strong acids but is readily destroyed by alkali, being hydrolyzed into its constituent pyrimidine and thiazole rings. It is adsorbed from solution by Fuller's earth. It was isolated as the chloride in crystalline form by Jansen and Donath in 1926. 0.002 mg of this preparation was stated to cure polyneuritis in pigeons. It was given the following empirical formula by Windaus and his associates—C₁₂H₁₈N₄OSCl₂. It was named *thiamin* because it contained sulfur, and at the time was the only known vitamin that did. Thiamin was synthesized by Williams in 1936. The following is the structural formula of thiamin hydrochloride.



Thiamin hydrochloride (C₁₂H₁₈N₄OSCl)

Distribution The richest sources of vitamin B₁, either as the pyrophosphate (cocarboxylase) or in the free form are, liver, kidney, lean pork, brewer's yeast, germ of wheat, oatmeal, rice polishings, bran, soy bean, peanuts, peas and beans. Smaller amounts are present in egg-yolk, milk, potatoes, corn meal, turnips, bananas and various nut kernels. The several factors of the B complex tend to show roughly parallel distributions, for example, foods which have a high content of thiamin are generally rich in other B factors. Thiamin is synthesized to a moderate degree in the intestinal tract.

The daily adult human requirement is around 0.6 mg per 1000 calories of diet. The requirement is increased during the actively growing period, in

pregnancy and lactation and, as already mentioned, on a high carbohydrate diet. The requirement is also raised by the increased metabolism of fever or hyperthyroidism, and by a low environmental temperature, but is reduced by raising the fat content of the diet. This effect of dietary fat is spoken of as its "sparing effect" upon the B₁ requirement. It does not appear, however, to be a true sparing action, for no significant difference is demonstrable between the thiamin content of the tissues of B₁-deficient animals on high and low fat diets, respectively. The experiments of McHenry and his associates indicate a relationship between the actions of choline, thiamin and fat metabolism. The action of thiamin is antagonized by its pyridine analogue *pyrithiamine*, and destroyed by a factor in raw, fresh fish (p. 743).

Effects caused by vitamin B₁ deficiency

(a) *Beri-beri—polyneuritis* Beri-beri is a disease which for centuries has been prevalent in rice-eating countries, e.g., Japan, China, India, Dutch Indies, Philippine Islands, etc. It is also occasionally seen in Labrador and Newfoundland and in young children upon deficient diets. The disease is characterized by inflammation of the peripheral nerves (polyneuritis) which leads to progressive paralysis of the limbs and sensory disturbances. There is also dilatation of the right heart. The disease occurs in two forms: one with edema (wet type) the other without (dry type). The first definite indication that beri-beri was of dietary origin was obtained (1885) by Takaki, a medical officer of the Japanese Navy who, by revising the diet of the sailors, practically eradicated the disease from the service. He thought, however, that the disease was due to protein deficiency and replaced a part of the polished rice of the diet by meat, milk, wheat and barley. The experimental investigation of beri-beri dates from 1890 when the Dutch physician Eijkman observed a disease in fowl at his laboratory in the Dutch West Indies which he believed was of the same nature as the human condition. The affected fowl had received a diet of polished rice, i.e., rice from which the pericarp (bran) and germ had been removed by milling (see fig. 55.6, p. 773). It was soon shown by others (Fraser and Stanton) that the condition in fowl and beri-beri in man could be cured by the addition of rice polishings to the diet or an extract prepared from them. The subsequent preparation by Funk of a crystalline substance of high anti-

neuritic potency from rice polishings has been mentioned

Polyneuritis is readily induced in pigeons by a diet composed exclusively of polished rice. Retraction of the head and paralysis of the limbs are outstanding features. B_1 -deficient diets cause analogous symptoms in dogs and rats (see fig 54.3). The oxygen consumption of the brain tissue is reduced. An accumulation of lactic acid in the hind-brain (Kinnorslev and Peters), liver, heart and muscles of birds occurs in the advanced stages of the disease and the lactic acid of the blood is



FIG 54.3 Upper photograph, dog, showing polyneuritis with marked paralysis of hind limbs as a result of a diet lacking in vitamin B. Lower photograph, the same animal cured by vitamin B given in the form of tomato juice (After Cowgill and Mendel)

raised. The head retraction is probably due to the increased concentration of lactic acid in the brain substance, the symptom is quickly abolished by the local injection of the vitamin. A high carbohydrate diet intensifies the effects of B_1 deficiency which fact taken together with the high lactic acid concentration just mentioned points to a fault in carbohydrate metabolism as the fundamental cause of the symptoms. Hyperglycemia and depletion of liver glycogen occur in polyneuritic pigeons. Though this may be due in part to inanition, for Drummond and Marnan observed it

in starved pigeons receiving adequate amounts of vitamin B_{12} , it is mainly a direct effect of the avitaminosis. The abnormally high glucose tolerance curves (i.e., reduced glucose tolerance) of rats showing signs of B_1 deficiency also furnish evidence of a disturbance of carbohydrate metabolism (Lepkowsky and associates). *In vitro* experiments indicate that an action of B_1 is to aid in the oxidation of carbohydrate in the brain through the lactic acid stage. For example, sliced brain tissue from an animal (pigeon or rat) suffering from vitamin B_1 deficiency does not consume the normal amount of oxygen when lactic acid is added to it, pyruvic acid also appears in the avitaminosis brain tissue but not in the normal brain. The accumulation of pyruvic acid inhibits in turn the removal of lactic acid. Failure in the disposal of both these intermediaries of carbohydrate metabolism therefore occurs in the avitaminosis brain. The addition of vitamin B_1 partially corrects these defects, the oxygen usage is increased and the production of pyruvic acid is reduced. The vitamin exerts a negligible effect upon the respiration of healthy brain tissue.

Vitamin B_1 occurs in the tissues both as the free base and as the pyrophosphate, which is known as *cocarcboxylase*. This latter form is present as a prosthetic group of the enzyme carboxylase which is required for the decarboxylation of pyruvic acid as well as of α -ketoglutarate and α -ketobutyrate.

(b) *Arrested growth* (fig 54.4) Osborne and Mendel found that "protein-free" milk supplied a substance necessary for the growth and well-being of rats. They concluded that this substance was identical with the antineuritic vitamin. It is now recognized that B_1 as well as other factors of the B complex influence growth. Loss of appetite, which leads to undernutrition, is, however, a contributory factor in the retardation of growth resulting from a deficiency of this vitamin. In infants the failure to gain at the normal rate though the diet appears to be adequate, is considered in some instances to be due to a low B_1 intake. Tisdall has reported that in a group of young children to whom a concentrate of the B complex was administered, the rate of weight gain over a period of seven months was 1.6 times that of a control group on the same diet but receiving no additional supply of the vitamin.

(c) *Loss of appetite and along of the gastrointestinal tract*. Loss of appetite is an early effect of vitamin B_1 deficiency, being evident some time

before the appearance of polyneuritis. It occurs in rats and dogs and in human beri-beri. It is probable that the anorexia is, in part at least, secondary to a relaxed state of the gastro-intestinal musculature and the reduction in motor activity of the gastro-intestinal tract which have been shown by Cowgill and associates to be important features of vitamin B₁ deficiency. McCarrison produced atony of the bowel and degeneration of the mucous membrane of the colon in monkeys by feeding diets lacking in the B₁ vitamin but of high carbohydrate content. It is not unlikely that in many diets which are assumed to be adequate, the

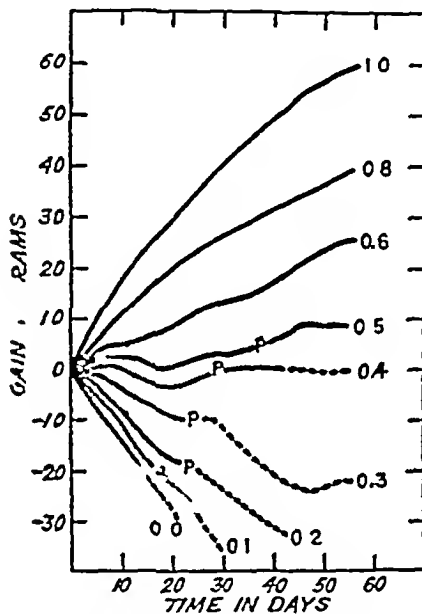


FIG 544 Average weight gain curves of rats on vitamin B-free diet plus daily supplement of ground whole wheat of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 and 1.0 grams. Negative controls (marked 0.0) received basal diet only. *P, point where, on the average, chronic symptoms of polyneuritis appeared. Lines broken from the points at which some individuals died (After Chase, from Sherman and Smith, *The Vitamins*.)

B₁ content is below the optimal level, and that gastro-intestinal abnormalities are not uncommonly the result of mild grades of deficiency of this vitamin, or other factors of the B complex. Yet recent work suggests that the effect of thiamin upon gastro-intestinal function may have been exaggerated and that other components of the B complex, notably, pantothenic acid, have a greater influence.

(d) *Bradycardia (in rats and pigeons)* Marked cardiac slowing was observed by Carter and Drury in pigeons fed upon polished rice, it was shown to

be due to increased tone of the vagus center, it is abolished by vagal section or atropinization. Drury, Harris and Maudsley observed an extreme degree of bradycardia in rats upon diets deficient in vitamin B₁ but in these animals it is not of vagal origin, it arises in the sinus node. The administration of material rich in vitamin B₁ restores the cardiac rate to normal within an hour or so. It has been shown by Birch and Harris that lactic acid accumulation in the cardiac tissue is associated with the phenomenon.

(e) *Personality changes and alcoholic psychoses* Thiamin deficiency of a degree insufficiently severe to cause beri-beri may result in psychic changes, such as depression, irritability, quarrelsomeness, anxiety or fearfulness. Alcoholic psychoses, e.g., *Wernicke's encephalopathy*,

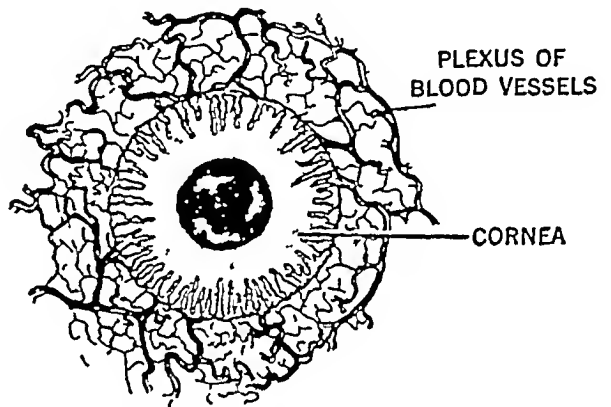
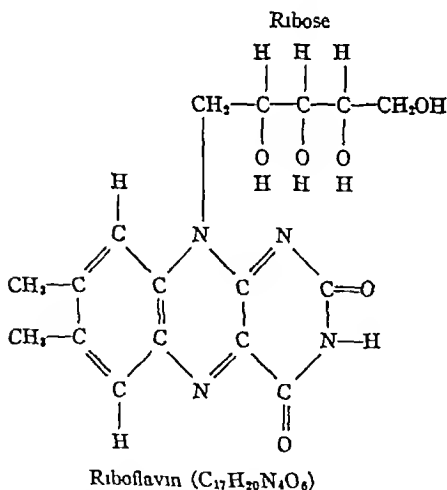


FIG 545 Showing engorgement of the limbic plexus and vascularization of the cornea in riboflavin deficiency. Note the loops of vessels of the limbic plexus penetrating more than half way to the pupil. White areas are corneal opacities.

ally, are believed to be due in many instances, in part at least, to a deficiency of the antineuritic vitamin. The chronic alcoholic receives a large proportion of his caloric requirement in the form of vitamin-free alcohol. Wernicke's encephalopathy is a condition marked by various neurological manifestations, paralysis of extrinsic ocular muscles, loss of memory, mental confusion, confabulation and coma. It occurs in chronic alcoholics and less commonly in other forms of severe undernutrition. Though thiamin deficiency is the chief cause it is not the sole one. Vitamin B₁ deficiency rather than a direct toxic effect of alcohol, is held responsible for the aggravation of preexisting cardiac disease and the precipitation of cardiac failure (if not for originating it) which is of such frequent occurrence in chronic alcoholism. In other words, the drunkard may show the cardiac features of beri-beri in varying degrees.

RIBOFLAVIN

Riboflavin^a belongs to a group of yellow fluorescent pigments called *flavins*. It may be present in the tissues as such or conjugated with protein as a flavoprotein. It is an orange-yellow substance which exhibits a greenish-yellow fluorescence in aqueous solution. It is moderately heat stable, especially in acid solution, but is rapidly destroyed by light. This factor is responsible for a part of the growth promoting property of the B complex. It is present in the tissues both free and as a prosthetic group in cytochrome C reductase, d-amino-acid oxidase and xanthine oxidase. A reduction of these enzymes has been demonstrated in the tissues of animals suffering from riboflavin deficiency. The flavins are soluble in aqueous media and for this reason are called lyochromes, as distinguished from the fat soluble pigments—the lipochromes. Flavins are widely distributed in animal and plant tissues. The chief sources of riboflavin are liver, kidney, milk, and the green leaves of vegetables. This vitamin has been synthesized artificially and moderate amounts are formed by the intestinal flora. It has the following formula



The daily adult human requirement of riboflavin is between 2 and 3 milligrams.

In man riboflavin deficiency has been recognized as the cause of certain well defined ocular lesions. In the milder grades of deficiency, examination with the slit lamp (ch 75) reveals congestion of the vessels of the limbic plexus at the periphery

^a Confusion in terminology sometimes arises owing to the fact that riboflavin is referred to by some as vitamin B₂ or as vitamin G.

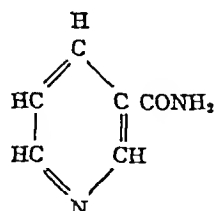
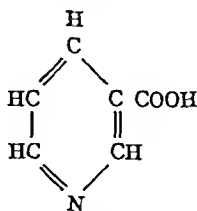
of the cornea and the invasion of the subepithelial layer of the cornea by loops of capillary vessels. When the deficiency is more severe the vascularization of the cornea is obvious to ordinary inspection, the eyes appear bloodshot (fig 54.5). Burning, itching and lachrymation are complained of. There is photophobia which may be so severe as to prevent the eyes from being opened. Interstitial keratitis may be a later development. Other effects of riboflavin deficiency are, inflammation and sores of the labial mucosa, and the development of fissures at the mucocutaneous borders, especially at the angles of the mouth (*cheilosis*), and smoothness with soreness of the tongue simulating the glossitis of pernicious anemia.

Animals on a diet lacking in this vitamin grow poorly, chicks develop a characteristic paralysis (curled toe) with myelin degeneration of the peripheral nerves, dermatitis, cataract, atrophy of the testes and loss of hair are produced in rats, and corneal opacities in dogs.

NICOTINIC ACID (NIACIN) AND NICOTINIC ACID AMIDE (NICOTINIC AMIDE OR NICOTINAMIDE)

(pellagra preventive factor)

Chemistry, metabolism and function The formula of nicotinic acid and nicotinic acid amide are shown below —



Nicotinic acid, though related to nicotine (from which it can be derived through oxidation by nitric acid or potassium permanganate), does not exhibit the action of nicotine upon the autonomic nervous system and is relatively non-toxic. It does, however, cause peripheral vasodilatation with flushing of the skin when injected intravenously. Nicotine absorbed in smoking cannot be converted into the vitamin. Niacin is excreted in the urine mainly as N'-methyl nicotinic amide. The vitamin is stable at high temperature either in solution or in the dry form.

Synthesis of nicotinic acid has been shown to occur in rats, sheep and chicks and is brought about largely by bacterial action in the intestine, though in the chick, at any rate, synthesis also appears to be a function of the body tissues

This vitamin acts principally as a coenzyme in tissue respiration. Co-enzyme I (diphosphopyridine nucleotide) (ch 32) and co-enzyme II (triphosphopyridine nucleotide) contain it as a prosthetic group. A reduction of the former enzyme has been demonstrated in the tissues of animals suffering from nicotinic acid deficiency. The concentration of these enzymes in the blood of normal human subjects, as well as of those suffering from pellagra, is raised by the administration of nicotinic acid. The function of this vitamin is closely associated with the metabolism of carbohydrates, the requirement being raised in diabetes when the consumption of carbohydrate or the insulin dosage is increased.

Sources and requirement Liver (especially pork liver), kidney, lean meats, brewer's yeast, rye, wheat germ, soybean and peanuts are among the richest sources of this vitamin. Corn meal and white flour contain less than 2 mg per 100 grams. The daily adult human requirement lies between 12 and 20 mg. This is about the amount contained in one serving of liver or a quart of milk.

Effects of nicotinic acid deficiency

Deficiency of nicotinic acid or its amide is the cause of the main symptoms of *pellagra* and of the condition in dogs known as "*black tongue*."

Pellagra The chief features of pellagra (= rough skin) are patches of dermatitis, redness, edema and soreness of the tongue together with intestinal ulceration, digestive disturbances and diarrhea. Nervous disorders (muscular weakness, tremor, paresthesias) may occur, and in the later stages melancholia, dementia or delirium. The skin lesions consist of redness, dryness and the formation of scales upon surfaces exposed to the sun's rays, e.g., the backs of the hands (or the insteps of persons who go barefoot), the neck, cheeks and bridge of the nose. A reddish pigment is excreted in the urine. Degeneration of the spinal tracts, especially of the posterior columns, and of nerve cells in brain and cord is not uncommonly seen post mortem.

Just as beri-beri is a disease affecting those who subsist mainly upon a diet of polished rice, so pellagra is a maize-eater's disease. The condition

results whether the maize is whole or has had the pericarp and germ removed. It is prevalent among the poor of the southern United States, of Spain, Italy and other European countries. It is rare in England, Canada and the northern United States. The dietary origin of pellagra was clearly demonstrated by Goldberger. He carried out an experiment upon twelve prisoners, who in return for a promise of pardon, volunteered to submit to a diet of cornmeal, cornstarch, rice, syrup, sweet potatoes and pork fat. At the end of six months pellagra was diagnosed in half the subjects. The other prisoners who received the ordinary institutional fare showed no signs of the disease. At first, Goldberger thought that the disease was due to protein (amino acid) deficiency but as a result of subsequent experiments upon animals he came to the conclusion that a vitamin deficiency was the principal causative factor. It soon became evident that autoclaved yeast, wheat germ and other substances rich in the vitamin B complex would prevent or cure pellagra. Yet it has been found more recently that deficiency of tryptophane is also an essential factor in the development of the disease. Zein, the main protein of maize, has a very low content of this amino acid. Rats upon a diet, adequate except for a lack of nicotinic acid, remain in good health unless tryptophane is also at a low level, when the characteristic signs of deficiency appear. Tryptophane appears to be a precursor from which this vitamin is synthesized by the intestinal bacteria. It can replace the vitamin in the diet of rats and when ingested by man and other mammals the excretion of nicotinic amide is increased. Another factor in the production of pellagra is the presence in maize of a principle which antagonizes the action of nicotinic acid (see p 743).

Most subjects of pellagra also suffer from deficiencies of other factors of the B complex, the administration of nicotinic acid alone being as a rule insufficient to effect a complete cure.

"*Black tongue*" in dogs is a condition analogous to human pellagra. Necrotic areas appear upon the tongue and the buccal mucosa, and a pellagra-like dermatitis of the scrotum develops in animals receiving a diet free from vitamin B₂.

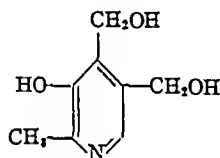
Until the last few years the factor in the B complex which prevents pellagra (P.P. factor) had eluded discovery. Funk as early as 1911 reported the presence of *nicotinic acid* in his extracts but its physiological significance was not realized. War-

burg recognized its presence in association with tissue respiratory enzymes and was aware of its fundamental importance in the oxidative processes of the cell. In 1937, Elvehjem and his colleagues isolated nicotinic acid and its amide from liver and discovered that it cured "black tongue" in dogs. Shortly afterwards it was used with outstanding success by Spies and his associates in the treatment of pellagra.

Certain mental and neurological manifestations, e.g., clouding of consciousness, rigidities and, in infants, unusual grasping and sucking reflexes have been attributed to nicotinic acid deficiency. Improvement has followed administration of the vitamin.

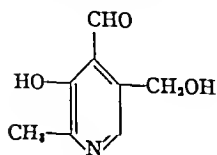
PYRIDOXINE

Pyridoxin or *adermin*, *Vitamin B₆* (antidermatitis factor). Rats upon a diet lacking in vitamin B₆ develop a skin disorder (rat dermatitis or acrodynia) characterized by redness, scaliness and loss of hair. It is now agreed that the antidermatitis effect of the B complex which has been recognized for years is due to this factor. Vitamin B₆ was synthesized by Harris and Folkers in 1939. It has the following structural formula:

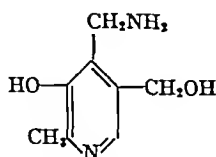


Pyridoxine (2-methyl-3-hydroxy-4,5-di[hydroxy-methyl]pyridine (C₈H₁₁O₃NCl))

Two natural derivatives of pyridoxine, *pyridoxal* and *pyridoxamine*, are more effective than pyridoxine itself. Both are found in animal and vegetable tissues.



Pyridoxal



Pyridoxamine

Pyridoxine serves as a coenzyme in the decarboxylation of tryptophan, arginine, glutamic acid and dihydroxyphenylalanine ("dopa") (p. 828), and in transamination reactions. It also appears to play a rôle in the metabolism of tryptophan and

there are indications that it is connected with the metabolism of unsaturated fatty acids.

Spies and Ashe have reported that certain symptoms in pellagrins and subjects of beri-beri, namely, extreme nervousness, insomnia, irritability, abdominal pain and difficulty in walking, which are not relieved by nicotinic acid, thiamin, or riboflavin, are abolished in dramatic fashion by pure B₆ (synthetic).

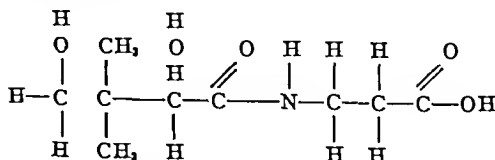
There are also indications that this vitamin is of importance in hemoglobin synthesis and in the manufacture of red cells. Dogs and pigs on a diet deficient in pyridoxin develop a hypochromic microcytic anemia which is quickly cured when the deficiency is corrected. Also, in rabbits rendered anemic by repeated bleedings, hemoglobin regeneration and erythrocyte production are hastened by the administration of pyridoxin. This vitamin is also necessary for the normal growth of rats; an assay method is based upon this fact.

The essential importance of B₆ in the life of many lower forms has been demonstrated. It is necessary for the growth of yeast, mosquito larvae and several types of microorganisms (*Streptococcus hemolyticus*, *Staphylococcus albus* and lactic acid bacillus).

The richest sources of pyridoxine are certain vegetable fats, milk, rice, wheat germ, yeast, legumes, meat and meat products, especially muscle and kidneys. Though in man no definite signs or symptoms attributable to pyridoxine deficiency alone are known, benefit from its use has been reported in Parkinsonism, muscular dystrophy and myasthenia gravis, but hopes that it would be of real value in these disorders have been on the whole disappointed. The adult human requirement is unknown but 15 mg daily has been suggested based upon animal experiments.

PANTOTHENIC ACID (CHICK ANTIDERMATITIS FACTOR)

This component of the B complex has the following formula:



Pantothenic acid (α γ -dihydroxy- β , β -dimethylbutyryl aminopropionic acid)

It was found by Williams and his associates in extracts of various plant and animal tissues and

was shown to stimulate the growth of yeast. Like other factors of the B complex already considered, pantothenic acid has been synthesized. Its identity with the factor, which for some time has been recognized as preventing dermatitis in chicks, is now established. Lesions of the spinal cord and eye symptoms are also produced in chicks by a deficiency of this factor. It is necessary in the diet of the hen for the hatchability of the eggs, and for egg-laying in pullets.

Pantothenic acid deficiency in rats causes a condition known as "spectacle eyes", this appearance is given by a hairless, inflamed ring of skin around the eyes. Renal and cardiac damage, dehydration and sometimes necrosis of the adrenal cortex are found. There is also an increased appetite for salt, which is probably the result of the adrenal defect. This vitamin has been termed an anti-gray hair factor since it prevents graying of the fur in black rats. Depigmentation of the feathers of fowl suffering from a deficiency of pantothenic acid has also been reported. These results upon animals cannot be transferred to man. No definite evidence of a factor which prevents the graying of human hair has yet been secured. Pantothenic acid is also necessary for the optimal growth of rats.

No definite nutritional defects in man due to pantothenic acid deficiency are known. The human requirement is also unknown but from animal experiments and the quantity excreted by the human kidneys it is probably between 3 and 5 mg.

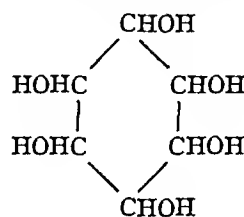
Pantothenic acid is very widely distributed throughout nature, being present in nearly all animals and vegetable tissue, from this ubiquity its name was derived. The richest sources are, liver, kidney, egg-yolk, wheat bran, brewer's yeast, broccoli, spinach, molasses and peanuts.

INOSITOL

Inositol (mouse anti-alopecia factor) $C_6H_6(OH)_6$, a carbohydrate (a hexahydroxycyclohexane) is found combined with phosphate, calcium and magnesium in cereals, the complex being known as *phyltin* (p. 767). It is also present in the lipid of soy bean, and in animal and plant tissues generally. In plants, it is present in greatest abundance in the leaves. Of animal tissues, liver, kidney, skeletal and heart muscle contain the largest amounts. This factor is necessary for the normal nutrition of the mouse and rat. When absent from the diet of young mice, growth is arrested, the hair

falls out, and fails to be restored, the animals becoming quite bald, dermatitis develops. Rats on an inositol-deficient diet show "spectacle eyes" and fatty livers. Its function in other species and in human nutrition is unknown, but the possibility of its being of benefit in certain skin lesions, e.g., psoriasis, has been suggested. It is also reported that inositol enhances the beneficial action of vitamin E upon the course of progressive muscular dystrophy, possibly by forming a complex with tocopherol in the intestinal tract which is necessary for normal creatine metabolism. The requirement of this vitamin is unknown.

Inositol has the following structure



Inositol $C_6H_6(OH)_6$

The function of inositol is associated in some way with the metabolism or transport of fat, for it is capable, like choline and lipocaine, of preventing the accumulation of liver fat.

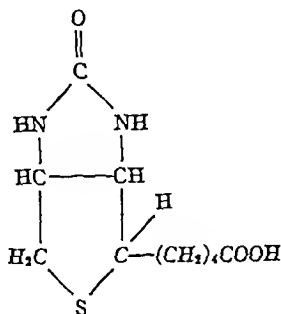
BIOTIN

Biotin (vitamin H). Many years ago (1901) Wildiers found that yeast required a principle for its growth in addition to the essential factors which were then known. He named it "*bios*". It was shown by Lucas more than twenty years later that "*bios*" was a mixture, and he separated it into two substances, bios I and bios II. Shortly afterwards bios I was identified as inositol and bios II shown to be a composite of at least two substances. One of these was isolated in pure form by Kogel and Tonnies and named biotin.

Biotin, as well as being necessary for the growth of yeast, is a growth stimulant or an essential requirement for certain bacteria and fungi. It has also been found to be an essential nutrient for animals (Boas) and man (Gyorgy). This vitamin appears to be essential for the synthesis of an enzyme which catalyzes the carboxylation of pyruvic acid to form oxaloacetic acid, which can then be converted by transamination to aspartic acid. It is also required for the reverse reaction—decarboxylation of oxaloacetic acid to form pyruvic acid. Oxaloacetic or aspartic acid can partially replace

biotin as a growth factor for certain micro-organisms (e.g., *Lactobacillus arabinosus*)

The chemical structure of biotin was determined by du Vigneaud and his associates and its synthesis effected by Harris and his colleagues. The structural formula of biotin (2' keto 3,4, imidazolido-tetrahydrothiophenovaleric acid) is shown below



Biotin ($C_{10}H_{16}O_2N_2S$)

Rats fed upon a diet of raw egg-white develop a severe dermatitis, alopecia, "spectacle eyes" and a spastic state which terminates in death. This condition, known as "egg-white" injury, is due to the presence in egg-white of an albumin called *avidin*, which combines in stoichiometric proportions with biotin in the intestine and renders it unavailable in nutrition. Biotin also occurs in this inactive bound form (*avidin-biotin*) in foods as well as in a free active form. Cooked egg-white does not produce ill effects, probably because the avidin is denatured and does not combine with the vitamin. Biotin is not liberated from its combination with avidin by the digestive, proteolytic enzymes, but separation can be brought about by oxidative procedures, e.g., treatment with hydrogen peroxide. Avidin appears to be associated in some way with the female reproductive functions. Stilbestrol administration followed by progesterone causes the production of this protein in the oviducts of immature chicks.

Biotin deficiency is produced experimentally in man when about 30 per cent of the total calories of the diet is furnished by dried egg-white. The symptoms consist of extreme pallor of the skin and mucous membranes and a seborrheic dermatitis, together with great lassitude, sleepiness, loss of appetite, pains in the muscles and precordial distress. These effects are almost immediately relieved by the parenteral administration of the vitamin.

The biotin requirement of man is around 200 micrograms per day. Like several other factors of the B complex, this vitamin is synthesized in some degree by the bacteria of the intestine.

Biotin is found in greatest amounts in liver, kidney, brewer's yeast and poultry. It is present in lower concentrations in egg-yolk, tomatoes and carrots. Certain types of carcinoma contain it in relatively large amounts. This finding has suggested the possibility that, by reducing the free biotin content of the diet, inhibition of cancerous growth might be brought about.

PARA-AMINOBENZOIC ACID (ANTI GRAY HAIR FACTOR)

Para-aminobenzoic acid This factor has the following formula



Para aminobenzoic acid (PABA $C_7H_7O_2N$)

It is essential for the growth of certain microorganisms and antagonizes the bacteriostatic action of the sulfonamides. The latter drug, according to the theory of Woods, competes with para-aminobenzoic acid (PABA), which it resembles chemically, for a place in some metabolic process or enzyme system in the bacterial cell necessary for its reproduction, thus the action of PABA is blocked. An excess of PABA prevents in turn the bacteriostatic action of the drug. The resistance to these bacteriostatic agents which is shown by *Staphylococcus aureus* and other microorganisms is due to the ability of such organisms to synthesize para-aminobenzoic acid. This factor is also an essential constituent of the diet for the normal growth of chicks, and in combination with inositol, prevents decoloration of the hair (achromotrichia) of rats. The effect upon the growth of chicks is probably an indirect one, namely, the stimulation of the growth of intestinal bacteria and the synthesis of other growth-promoting factors. The graying of the hair of black rats on a synthetic diet is prevented by the addition of PABA. As with other vitamins, no relationship between PABA and the graying of human hair has been established. Graying of hair in young children on grossly deficient diets and suffering from severe malnutrition has been reported but the dietary

factor or factors which are lacking have not been identified

An important clinical development is the discovery of the beneficial effect of this vitamin in the treatment of louse borne typhus and Rocky Mountain spotted fever

uronic acid It has been appropriately named *ascorbic acid* ($C_6H_8O_6$)⁷

Ascorbic acid was isolated by Szent-Gyorgyi in 1928 from oranges, lemons and cabbages and from the suprarenal cortex It was shown to have a high antiscorbutic potency Almost simultane-

TABLE 74
The content of some typical foods in the vitamin B complex
(From Elvehjem, slightly modified)

FOODS*	THIAMINE†	RIBOFLAVIN†	NIACIN†	PANTOTHENIC ACID††	VITAMIN B ₆ ††	BIOTIN††	FOLIC ACID††
Apples	0 04	0 02	0 2	0 05	0 03		
Bananas	0 09	0 06	0 6	0 18	0 30		0 01
Bread							
White (unfortified)	0 08	0 13	0 8	0 40	0 20		
White (fortified)	0 24	0 15	2 2	0 40	0 20		
Cabbage	0 07	0 06	0 3	0 08	0 29		0 01
Carrots	0 07	0 06	0 5	0 24	0 19	0 002	0 01
Cheese	0 04	0 50	0 1	0 35	0 20	0 002	
Cornmeal, degerminated	0 15	0 06	0 9		0 25		0 02
Eggs, whole fresh	0 12	0 34	0 1	2 70		0 025	0 01
Meat							
Beef	0 12	0 15	5 2	1 10	0 40	0 004	0 02
Pork loin	1 04	0 20	4 4	1 50	0 60	0 005	0 01
Poultry, chicken or turkey	0 10	0 18	8 0	0 90	0 20	0 01	
Liver, pork or beef	0 27	2 80	16 1	5 20	0 80	0 1	0 08
Milk, whole fluid	0 04	0 17	0 1	0 30	0 07	0 005	
Oatmeal	0 65	0 14	1 1	1 30	0 25		0 03
Oranges	0 08	0 03	0 2	0 12			0 01
Peas, fresh	0 36	0 18	2 1	0 60	0 05	0 002	0 03
Peanuts, roasted	0 30	0 16	16 2	2 5	0 30		
Potatoes	0 11	0 04	1 2	0 40	0 16		0 01
Spinach	0 12	0 24	0 7	0 7	0 08	0 002	0 18
Tomatoes	0 06	0 04	0 6	0 37	0 07	0 002	0 01
Turnips	0 06	0 06	0 5	0 25	0 10	0 002	
Yeast, Brewers' dry	9 69	5 45	36 2	20 00	2 90	0 2	0 7
Wheat, whole	0 56	0 12	5 6	1 30	0 40	0 005	0 05

* Edible portion

† Values are given in milligrams per hundred grams

†† Values for pantothenic acid, pyridoxine (vitamin B₆), biotin and folic acid are based on data from only a limited number of samples Some of the values may be low because of incomplete liberation of the vitamin This is especially true in the case of pantothenic acid

CHOLINE

Choline (see p 696), *folic acid* and B₁₂ (ch 9) are dealt with elsewhere

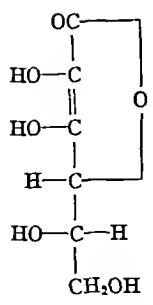
VITAMIN C (ANTISCORBUTIC)

Properties

Vitamin C has been shown to be a relatively simple chemical substance closely allied to hex-

ously King and Waugh obtained crystals of ascorbic acid from concentrates of lemon juice which were capable in 0 05 mg daily doses of protecting a guinea pig from scurvy It was later synthesized by Hirst and associates who gave it the following structural formula

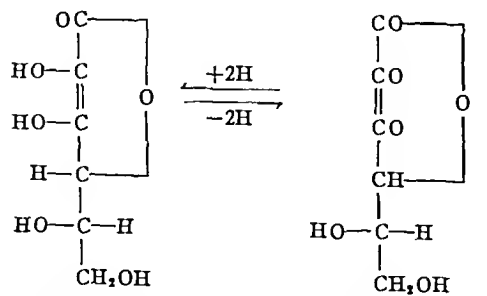
⁷ This is also referred to as *cevitamic acid*, a term recently adopted by the American Council on Pharmacy and Chemistry



1 Ascorbic acid—Vitamin C

Ascorbic acid has a high reducing power, is soluble in water and alcohol, and is readily destroyed by heat, sunlight and oxidizing agents. Prolonged boiling, drying, aging or storage of foods reduces or abolishes their antiscorbutic properties. Copper catalyzes the oxidative process, so that the vitamin is destroyed much more readily when foods are cooked in a copper utensil than when glass, enamel or aluminum vessels are employed. Nevertheless, boiling milk, even in an enamel vessel, for five minutes destroys 20 per cent of its vitamin C content. Heating to 60°C for thirty minutes (pasteurization) in a copper vessel destroys 80 to 90 per cent of its vitamin C value, but only a negligible loss occurs if the pasteurization is carried out in an aluminum or enamel container. In general, slow cooking, e.g., stewing, is more destructive to the vitamin than more rapid cooking even at a higher temperature. There is said to be little loss of vitamin C in the commercial canning of fruit.

Ascorbic acid is readily oxidized to dehydroascorbic acid and is as readily reduced again as shown below. This property of reversible oxida-



1 Ascorbic acid—Vitamin C

Dehydroascorbic acid

There is ample evidence that it serves such a function in the plant, but evidence for its playing this rôle in animal tissues is less certain. The respiratory activity of scorbutic tissue is but little reduced, nor does an increase in oxygen consumption result from the addition of ascorbic acid to such tissue. Oxidation of ascorbic acid to dehydroascorbic acid is thought to occur in the liver and to be dependent upon a special enzyme system requiring the presence of traces of copper.

The normal metabolism of the amino acids tyrosine and phenylalanine appear to be dependent upon the action of ascorbic acid. It has been observed by Levine and his associates that premature infants on a diet high in protein and deficient in vitamin C excrete parahydroxyphenyllactic and parahydroxypyruvic acids in the urine. The metabolic abnormality was exaggerated by feeding tyrosine or phenylalanine and corrected by the administration of vitamin C.

The distribution of vitamin C

The richest sources of this vitamin (as either ascorbic acid or as dehydroascorbic acid) are the citrus fruits (oranges, lemons, grapefruit and limes), cabbage, swedes and turnips, tomatoes, spinach, green and red peppers. Fresh meat, cow's milk and other animal foods are very poor sources of vitamin C. Its concentration is several times greater in human milk than in cow's milk. The vitamin is stored to a very limited extent in the body. It is in highest concentration in the adrenal cortex (ch 59). Of other tissues, the crystalline lens, the corpus luteum and the pituitary gland contain the largest amounts. The average concentration in human blood, when the vitamin C intake is adequate, is around 0.5 mg per 100 cc. From 30 to 50 mg are excreted daily in the urine at ordinary intake levels.

The synthesis of vitamin C

The precursor from which vitamin C is synthesized in nature is unknown, but glycuronic and galacturonic acids are likely possibilities. There is good evidence that glucose is a precursor. It is synthesized by germinating seeds and in the growing sprouts and tips of plants which contain it in relatively high concentration, whereas it is absent, or nearly so, from dried seeds and the less actively growing plant tissues. Ascorbic acid is formed by the chick embryo and apparently by the adult fowl, the rat, mouse and dog, but not by

tion and reduction appears to be closely bound up with its function in the tissues, namely, the transport of hydrogen

the guinea pig, the monkey or man. The particular tissue responsible for the process is unknown, it is not dependent, entirely at any rate, upon the adrenals since adrenalectomized rats and dogs can be maintained upon a diet free from vitamin C. This vitamin is now synthesized on a commercial scale from glucose, sorbitol being an intermediate product.

The effects of vitamin C deficiency

Scurvy (scorbutus) In man, the earliest sign of vitamin C deficiency is keratosis and enlargement of the hair-follicles. A little later increased vascularity of the follicles is seen and minute hemorrhages occur into them. These changes become evident as an eruption of small red papules. The essential pathological change in scurvy is weakening of the pericapillary sheath (p. 310) or possibly of the intercellular cement of the capillary wall itself.⁸ The abnormal fragility of the capillary walls leads to hemorrhage from various structures, e.g., the gums and the mucous membranes of the mouth and gastro-intestinal tract, skin, subcutaneous tissues, muscles and subperiosteal tissues. Redness, swelling, ulceration and, in severe cases, gangrene of the gums result. Some of the main features of the condition are anemia, small cutaneous hemorrhages (petechiae), pains in the bones and tender swellings, due to subperiosteal or muscular hemorrhages, separation of the epiphyses, especially in young children, great weakness and emaciation. X-ray examination of the scorbutic bone shows a white line running down the outside of the shaft near its ends (scorbutic white line) which is not seen in normal bone. In guinea pigs, the clotting time of the blood is prolonged and the platelets and red cells are reduced. There is a progressive reduction in the ascorbic acid content of the adrenal cortex in guinea pigs upon a scorbutic diet, the addition of orange juice to the diet restores the normal content.

In the past, scurvy as an adult disease occurred most frequently upon sea voyages, since fresh food for perhaps months at a time was lacking from the diet. For a similar reason it occurred during military campaigns and exploration parties, or in the general population in times of famine. During world war I it made its appearance among members of expeditions in the East, and among civilians in some of the warring nations.

It may also develop in artificially fed infants (*Barlow's disease*). According to Drummond the vitamin C content of the diet of adults is frequently, even at ordinary times, only a little above the level at which scurvy appears. Such grades of vitamin C deficiency may, however, be the cause of anemia, adequate supplies of the vitamin being required apparently for the normal functioning of erythropoietic tissue. Other manifestations of vitamin C deficiency short of scurvy are, loss of appetite, increased susceptibility to infection, edema of the gums, loss of weight and muscular weakness, and diminished capacity to metabolize tyrosine and phenylalanine.

One of the earliest records of the cure of scurvy by the administration of a substance rich in vitamin C is that describing an episode of Jacques Cartier's second voyage to Canada (1535). A number of the explorer's men had died from scurvy and most of those remaining were dangerously ill. These were cured by a drink prepared by a friendly Indian from the leaves and bark of an evergreen tree (probably the spruce).

About the middle of the eighteenth century (1747) Lind, a British naval surgeon, carried out an interesting and entirely admirable clinical research, the results of which showed conclusively the antiscorbutic value of oranges and lemons.

Lind, as he says, "took 12 men in the scurvy on board the 'Salisbury' at sea. Their cases were as similar as I could have them. They all had putrid gums, the spots and the lassitude with weakness of their knees." He grouped them into six pairs. To one pair he gave daily a quart of cider, to another pair elixir vitriol, to the third pair vinegar and to the fourth sea water and an electuary composed of garlic, mustard seed, balsam of Peru and myrrh, together with acidulated barley water as a drink. The fifth pair each had a lemon and two oranges daily. "The consequence was," he records "that the most sudden and visible good were perceived from the use of the oranges and lemons, one of those which had taken them being at the end of six days fit for duty. The other was the best recovered of any in his condition and was appointed nurse to the rest of the sick." He also noted some good effect from the cider but none from the other articles. Lind recommended that concentrated lemon juice be rationed to the navy. The admiralty did not follow this advice until 1795, after which the sobriquet "Lime juicer" was given by other nations to the British sailor.

The experimental investigation of scurvy may be taken to date from the work of the Danish scientists Holst and Frohlich (1907 to 1912) who

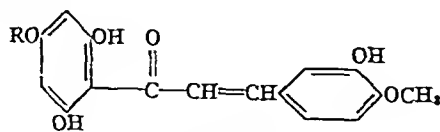
⁸ The capillary resistance test is described on page 122.

found that the condition developed in guinea pigs kept upon a diet completely lacking in green vegetables, e.g., one constituted solely of cereals. It is possible to induce scurvy only in those species which are incapable of synthesizing the vitamin, namely, in the primates and the guinea-pig.

The vitamin C requirement Harris and his colleagues have devised a method for determining the vitamin C requirement based upon the urinary excretion of ascorbic acid after a known amount has been given by mouth. Vitamin C is a threshold substance, i.e., it is reabsorbed by the renal tubules and does not appear in the urine if its concentration in the blood falls below a certain level, which depends in turn upon the requirement or degree of "saturation" of the tissues. If the subject's tissues are "saturated" with vitamin C a large part of the administered dose can be recovered from the urine, if the tissues are "unsaturated" more is retained, the proportion varying with the degree of unsaturation. The vitamin C content of the urine is determined by titration with 2,6-dichlorophenolindophenol. The daily requirement of vitamin C for the adult is around 75 mg,^{*} but it varies considerably. Infections, rheumatic fever and certain other conditions tend to deplete the vitamin stores and therefore increase the amount which must be provided in the diet. Infants require much more in relation to their body weight than do adults. The vitamin stores with which the baby comes into the world are depleted within the first few days and, especially if fed upon cow's milk, it should receive extra supplies in the form of orange or tomato juice.

Vitamin P (permeability vitamin) Szent-Györgyi and his associates obtained a crystalline substance from lemon juice and Hungarian red peppers which they claimed controlled vascular permeability. This material, which they called "citrin" or vitamin P, was found later to consist of two vegetable dyes (flavanones), *hesperidin* and *eriodictyol glucoside* (demethylated hesperidin). They stated that certain hemorrhagic diseases

associated with increased permeability or fragility of the capillary wall were cured by citrin, by lemon juice or by extracts of red pepper but not by ascorbic acid. They also found that the survival time of guinea-pigs upon a scurvy-producing diet was prolonged by vitamin P and that the hemorrhages were less pronounced in those that had received it than in control animals upon the scorbutic diet alone. It was therefore postulated that scurvy was due to a deficiency of both vitamin C and so-called vitamin P. After several conflicting reports concerning the latter, its existence appears to have been finally established. Results similar in nature to those just mentioned have been reported by Scarborough. Benefit followed the administration of "vitamin" P in cases of subcutaneous hemorrhages which failed to respond to vitamin C. It has not been established, however, that "vitamin" P is an essential dietary constituent and, therefore, should not be classified with the vitamins. Other substances such as *rutin* (quercetin-3 rhamnoglucoside) have a similar effect upon the capillaries. Griffiths and his associates found that this latter flavonoid prevented the increase in capillary fragility caused by the intraperitoneal administration of radon ointment in rats and in stimulating the reparative processes following x-ray burns. Clark and his associates in experiments with guinea pigs have also shown that rutin gives a high degree of protection against Roentgen irradiation sufficiently intense to cause widespread subcutaneous hemorrhages, whereas vitamin C is ineffective. The mortality in the treated animals was about half of that of the controls. The distribution in food stuffs of so-called vitamin P is closely similar to that of vitamin C. Lemon and orange peels are especially rich sources, but the vitamin is absent from milk, liver, kidney and other animal foods. It is in higher concentration in the peel of citrus fruits than in the juice or pulp. The physiologically active principle of this substance is believed to be *hesperidin chalcone*. Its formula is



Hesperidin chalcone

R is a sugar group

^{*} This figure is given by the U. S. National Research Council Technical Commission on Nutrition (1943), but is probably unnecessarily generous. More recent observations indicate that 30 mg is more nearly correct. As little as 10 mg has been found to provide protection against definite scorbutic manifestations.

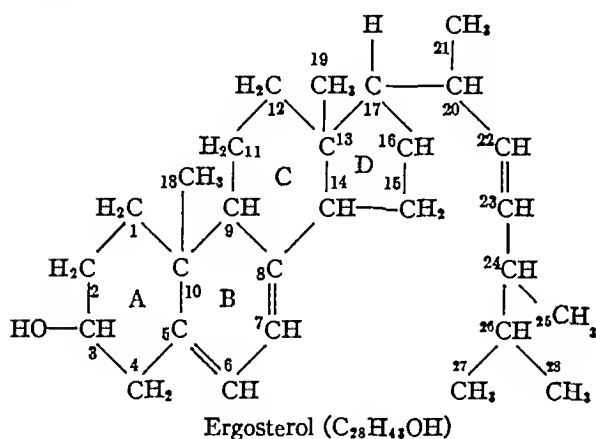
CHAPTER 55

THE VITAMINS (*Continued*)

VITAMIN D (ANTIRACHITIC)

Properties Vitamin D is soluble in fats, oils, ether and alcohol but insoluble in water. It is very stable to heat and to oxidation. It shows an absorption band between 260 and 270 $m\mu$ (maximum at 265). The antirachitic vitamin belongs to the class of substances known as sterols or solid alcohols. Among these are *cholesterol* ($C_{27}H_{46}O$) of animal tissues, *phytosterol* of plants, and *ergosterol* found mainly in fungi (yeast, mushrooms, ergot, etc.). *Ergosterol* was first obtained from mushrooms in 1811.

The following is the structural formula for ergosterol.



ULTRAVIOLET IRRADIATION AND VITAMIN D

The search for the antirachitic vitamin makes one of the most interesting stories of modern biological science. As early as 1822, Trousseau recommended cod liver oil for the treatment of rickets, and in 1890 Palm suggested that sunlight possessed an antirachitic action. In 1919, Huldchinsky successfully employed the ultraviolet rays from a mercury vapor lamp for the cure of rickets. Hess and Unger demonstrated in 1921 the antirachitic effect of sunlight. By this time Mellanby had published his results proving the antirachitic value of cod-liver oil, egg yolk and butter (see p. 766).¹ It was not long before intensive research by several workers furnished the information necessary to link together the antirachitic effects of these and other articles of diet with those of radiant energy. First it was found that when rats on a vitamin D defi-

cient diet were placed in glass jars which had been irradiated by a mercury vapor lamp they thrived as well as if they had been directly irradiated (Hume and Smith). The beneficial effect, as it turned out later, was due to the animals having eaten the irradiated sawdust in the jars. It was also discovered (Goldblatt and Soames) that the livers of rats which had received no vitamin D in their diet but had been irradiated with ultraviolet light acquired antirachitic properties similar to those of cod-liver oil. The livers of nonirradiated rats were quite ineffective. Shortly after this Steenbock and Black and Hess and Weinstock, independently, showed that certain fat-containing foods which possessed no power to cure rickets acquired this power upon artificial irradiation. It was then discovered by these two groups of observers and at the same time by Rosenheim and Webster that cholesterol of animal tissues (skin, brain, etc.) and sterols of vegetable foods gained antirachitic properties upon irradiation. Cholesterol, even after careful purification, and though quite inactive if untreated, became powerfully antirachitic upon irradiation. It was thought at first that the antirachitic property which cholesterol acquired upon irradiation with ultraviolet light was due to contamination with minute amounts of ergosterol. But it was shown later by Waddell and others that, though irradiated cholesterol was not antirachitic, a sterol in the skin closely associated with it, namely 7-dehydrocholesterol, was responsible, when irradiated, for the antirachitic action.

Ergosterol and 7-dehydrocholesterol are, therefore, the precursors of two vitamins D and are termed *provitamins*. The vitamin resulting from the irradiation of ergosterol is called vitamin D_2 , and that derived from 7-dehydrocholesterol is known as vitamin D_3 . There are a number of provitamins D which upon irradiation are converted to the corresponding vitamins, but only the two mentioned (D_2 and D_3) are of medical importance or interest. There is no vitamin D_1 .

Ergosterol, obtained from yeast, when irradiated yields a group of substances—*lumisterol*, *tachysterol*,² *calciferol*. During irradiation these products appear in the order given, but their proportions in

¹ Bland-Sutton used cod-liver oil with success in preventing rickets in lion cubs at the London Zoo a number of years earlier.

² Dihydrotachysterol is prepared through reduction by sodium and alcohol from an ester of tachysterol.

the mixture at any moment depend upon the intensity and duration of the irradiation. Calciferol is not the end result, but at a certain stage the amount of calciferol formed is at a maximum beyond which it undergoes decomposition, *toxissterol* and *supersterol* are produced. The first of these products of over-irradiation is highly toxic, having a very pronounced effect upon calcium metabolism. Calciferol (or *viosterol*), which has the greatest antirachitic effect and comparatively little toxic action, is regarded as the pure vitamin D₂, it can be isolated in crystalline form from the other products. A daily dose of 0.0001 mg of irradiated ergosterol is antirachitic for the rat, non-irradiated ergosterol is quite inactive. 0.025 micrograms (0.025 γ) of calciferol daily will prevent the development of rickets in a rat receiving a rickets-producing diet. The photochemical change in-

Dust, smoke or water vapor in the atmosphere being opaque to the shorter waves markedly reduce the antirachitic effect of sunshine. Ordinary window glass filters out all rays shorter than 320 m μ . Certain specially prepared types of glass are transparent to a proportion of the effective rays but sunshine transmitted through ordinary glass possesses no antirachitic action.

The antirachitic rays are incapable of penetrating the skin to any considerable extent beyond a depth of about half a millimeter or so. About 80 per cent of the rays with wave lengths between 250 and 300 m μ are absorbed or reflected from the corneous layer, the remainder are absorbed by the Malpighian layer and the corium. The blood in the capillaries of the corium acts as an effective filter, none of the ultraviolet rays penetrating beyond.

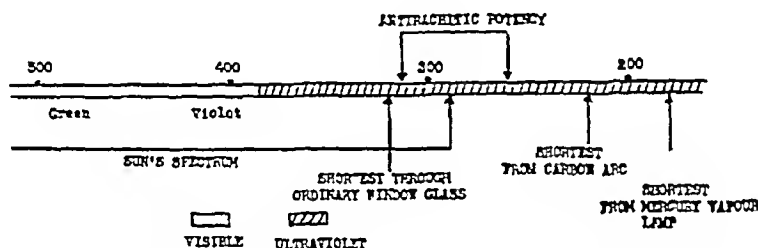


FIG. 55.1 Scheme of wave lengths of spectrum, showing the range of those possessing antirachitic properties. Figures refer to millimicrons. The hatched area indicates the ultraviolet part of the spectrum. (Redrawn and modified from Blunt and Cowan.)

involved in the activation of ergosterol is simply one of intramolecular rearrangement, nonirradiated ergosterol being isomeric with calciferol.

The ultraviolet rays effective in activating the provitamins are those which they absorb, namely, those with wave lengths between 250 and 313 m μ . The maximum effect occurs at a wave length of 281 m μ . Rays having wave lengths within this range falling upon the body surface are also effective in the prevention or cure of rickets. The antirachitic action common to irradiated foods and to direct irradiation of the body surface thus receives an essentially simple explanation.

The shortest rays from the sun which reach most localities of the earth have a wave length of about 290 m μ , whereas those from artificial sources such as the carbon arc and mercury vapor lamp are around 220 and 180, respectively (fig. 55.1)

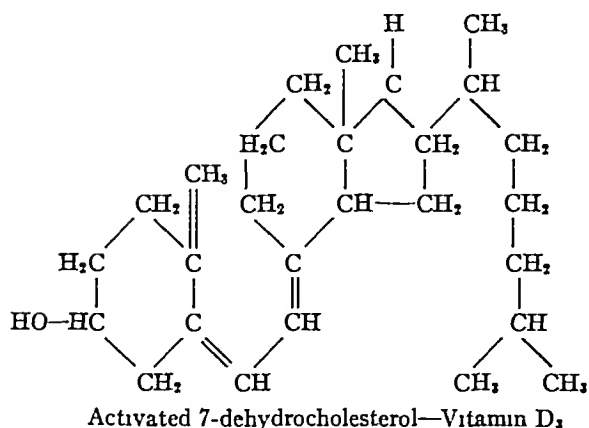
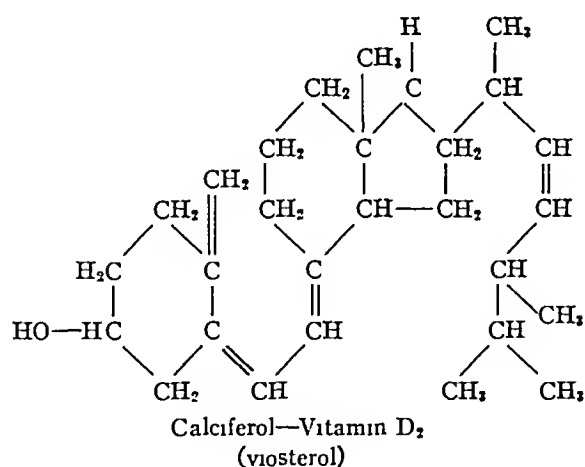
and is referred to briefly as AT 10 (antitetany compound No. 10). It is used therapeutically to raise the blood calcium in tetany. It has a relatively low antirachitic action, and its toxicity is higher than that of vitamin D.

A very interesting observation made by Hou upon birds may be mentioned in this connection. He found that the preen gland (glandula uropygialis) serves an antirachitic function. This gland, which is situated in the region of the tail feathers, secretes an oily substance. During preening the secretion is distributed over the body feathers where it is activated by the rays of the sun. The irradiated material is either absorbed through the skin or obtained from the ingestion of feather particles in subsequent preening operations. Rickets is said to follow extirpation of the gland and is not cured by irradiation of the general body surface. Fur-bearing animals probably obtain supplies of vitamin D in a similar manner, though the oily material is secreted not by a single structure but by glands distributed diffusely throughout the skin. The material which exudes becomes spread over the hairs and after irradiation either enters the body again through the skin or is acquired through licking the fur.

Calciferol, or vitamin D₂ as it is now called, and activated 7-dehydrocholesterol or vitamin D₃ are the two antirachitic vitamins of medical interest. There is no D₁, the original vitamin produced

by irradiation being a mixture of calciferol and lumisterol. The vitamin D of cod liver oil is mainly D_3 , whereas, in tuna liver oil vitamin D_2 predominates. The antirachitic potency of these two vitamins is not the same for all species. Vitamin D_3 is more potent (35 times) and D_2 for chicks, whereas D_2 is more effective in the rat, that is, a quantity of D_2 per gram of body weight which is adequate to prevent or cure rickets in the rat, is quite inadequate for the chick. In infants there appears to be little difference between the actions of the two vitamins. The vitamin D produced by irradiation of the skin or of milk is 7-dehydrocholesterol, that produced by the irradiation of yeast and which in oily solution has been marketed under various trade names (e.g., viosterol) is irradiated ergosterol (of yeast), i.e., calciferol or vitamin D_2 . Cows fed upon irradiated yeast simply transfer the vitamin to the milk which therefore owes its antirachitic properties to a vitamin different from that in irradiated milk.

The formulae for calciferol and activated 7-dehydrocholesterol are shown below. The formula for calciferol differs from that of ergosterol in the



breaking of the ring between carbon atoms 9 and 10, in having four double bonds instead of three and a CH_2 instead of a CH_3 group attached to the carbon at CH_{10} , i.e., at C_{18} . Activated 7-dehydrocholesterol has only three double bonds, lacking the one between C_{22} and C_{23} .

MAIN FOOD SOURCES OF VITAMIN D

Halibut-liver oil is one of the richest natural sources of vitamin D. Cod-liver oil and the liver oils of bony fishes generally are other rich sources. Cod-liver oil, however, contains only 100 international units (p. 774) of vitamin D per gram as compared with 1200 units per gram in halibut liver oil. Of all fish oils that from the liver of the bluefin tuna contains the greatest amount of vitamin D (up to 40,000 international units per gram). The liver oils of cartilaginous fishes are relatively poor in the antirachitic vitamin. Mammalian liver which is rich in vitamin A is poorly supplied with vitamin D. The cod and other larger fish receive their supplies of vitamin D, in part at least, from the bodies of the smaller fish upon which they feed. These in turn acquire it in a way similar to that in which they obtain vitamin A, i.e., from small plant-eating animal forms (zooplankton, p. 745) which are capable of synthesizing the vitamin. Bills concludes that the cod can also synthesize the vitamin. He does not believe it possible that the large stores of vitamin D in the liver of the cod can be derived solely from the food. But that it is produced by the transformation of the provitamin in the fish's skin by radiant energy seems to be out of the question, the cod being a deep-sea fish. It is possible that the tissues of fish contain an enzyme system capable of converting a provitamin to vitamin D. Another hypothesis which has been proposed is that the ultraviolet radiations (190 to 250 $m\mu$) which some believe are generated by metabolic processes in growing cells—the so-called mitogenetic or Gurwitsch rays—activate the provitamin.

Other animal sources of vitamin D are egg yolk, butter, cream and milk. The antirachitic potency of these dairy products is, as a rule, quite low and depends upon the vitamin D content of the diet and upon the extent to which the animal has been exposed to sunshine. Thus, summer butter is likely to have a higher antirachitic potency than butter produced in winter. Pasteurization does not lower the antirachitic property of milk. The egg yolk of irradiated hens or of hens receiving irradiated

ergosterol has a high vitamin D potency. Small quantities of vitamin D are present in beef fat but not in lard.

Vegetable foods are, as a rule, very poor sources of this vitamin. It is absent from most vegetable oils (cottonseed, or maize oil) unless these have been artificially irradiated. Green vegetables and fruits contain insignificant amounts. Yeast has a high content of ergosterol and when irradiated acquires a high degree of antirachitic potency. The milk of cows fed upon irradiated yeast or milk after direct irradiation is also potentially antirachitic.

An ordinary mixed diet is usually poorer in vitamin D than in any other vitamin, for this reason the diet of young children should be reinforced by this vitamin in some concentrated form. Adults probably receive all the vitamin D required in a well-balanced diet.

THE MODE OF ACTION OF VITAMIN D

Vitamins D exert their effect on calcium and phosphorus metabolism, but the exact details of this action have not been fully elucidated. It now seems certain, however, that vitamin D facilitates the absorption of calcium from the intestine. A negative calcium balance becomes positive under its influence. There is less certainty about the effect upon the absorption of phosphorus, though it can be said that the excretion of this element is reduced by the administration of vitamin D. Nevertheless, it appears that the principal action of the vitamin is probably the promotion of calcium and phosphorus deposition in the bones. Experiments with dogs given radioactive phosphorus support this conclusion. Moreover, while the blood phosphorus can be elevated by injections of glycerophosphate, and the blood calcium increased by dihydrotachysterol (AT 10), little or no effect upon the course of rickets is observed.

THE EFFECTS OF VITAMIN D DEFICIENCY

Vitamin D is indispensable to the normal calcification of bone. Its absence from the diet is followed by the development of rickets in infants or of osteomalacia in adults.

In infantile rickets (rachitis)

The fundamental feature of this disease is a disturbance of calcium-phosphorus metabolism with consequent defective ossification and the development of various deformities, e.g., *kroch knees*, *bow legs*, *enlargement of the epiphyses*,

spiral curvature, (*scoliosis*), *malformation of the chest*, *contracted joints*, *soft depressible areas in the parietal bones*, (*craniotables*) and the development of *bones of the temporal bones* (see fig. 55.2). The natural curvatures of the long bones tend to become exaggerated. The enlargements of the costosternal junctions—"beading of the ribs"—causes a series of small swellings on either side of the thorax which is referred to as the "*rachitic rosary*." Dentition is usually delayed. Sweating of the scalp is common.

CHANGES IN BONE STRUCTURE

The bones are relatively soft and pliable. *Defective calcification of the growing bone and compensation by hypertrophy of the epiphyseal cartilages* are the essential pathological changes. A section of the epiphyseal junction of a normal growing bone shows the following zones in order from the free ends of the bone toward the shaft.

- (a) A layer of resting *epiphyseal cartilage*.
- (b) A layer of *proliferating cartilage* in which the cells are arranged in regular columns paralleling the long axis of the bone.
- (c) A zone of *preparatory calcification*. The cells are still arranged in columns but deposition of calcium phosphate has occurred in the surrounding cartilaginous matrix.
- (d) A zone of *newly formed bone trabeculae*—the spongiosa—produced through the invasion of the foregoing zone by osteoblasts derived from the periosteum. Marrow tissue fills in the spaces between the trabeculae.

In a *rachitic bone* the layer of proliferating cartilage is greatly enlarged, being sometimes ten times its normal depth and the cells no longer show their regular columnar arrangement. This layer, instead of being sharply demarcated from the zone of preparatory calcification, sends finger-like cellular protrusions into the latter which is almost free from mineral deposit. The trabeculae are malformed and have lost their regularity of pattern. They are composed of *osseoid tissue*, i.e., a tissue very poor in or devoid of calcium and phosphorus. The cortex of the bone also may be partially replaced by osseoid tissue the extent to which this occurs varying with the severity of the disease.

On X-ray examination, the osseous abnormalities in a well-developed case of rickets are clearly evident. The entire shadow cast by the bone is less dense than normally and the ends of the bone are not sharply defined but have a "woolly" or "moth-eaten" appearance (fig. 55.3). The articular ends also frequently show a concave (cupped) rather than a convex or straight contour. This

appearance is due to the lack, or irregular distribution, of mineral in the zone of preparatory calcification. The latter normally presents a clearly defined almost straight band next to the unmineralized layer of proliferating cartilage.

Chemical analysis of the bones reveals a low mineral (Ca and P) content and a relative increase in organic matter and water. The ratio of ash to organic matter in normal, dry, fat-free bone is 3 to 2 whereas in rachitic bone it may be 1 to 2 or even 1 to 3. The normal Ca/P ratio remains unaltered (see also p 767).

The inorganic phosphorus of the blood in rickets is lower than normal, i.e., below 3.0 mg per 100 cc (it may be as low as 1.0 mg). This is an early sign. The serum calcium is usually normal unless



FIG 552 Child aged 5 years with leg deformity caused by rickets (After E. Mellanby)

the condition is complicated by tetany (p 849). The plasma phosphatase is elevated several fold (Kay) above the normal value. In well-marked rickets the *intestinal contents* tend to become less acid in reaction, the stools, which normally are slightly acid, neutral or slightly alkaline become definitely alkaline. Mineral retention is reduced in the disease, i.e., the calcium and phosphorus balances (p 862) show smaller positive values than the normal, or are negative.

THE RELATION OF CERTAIN FACTORS TO THE OCCURRENCE OF RICKETS

(1) *Diet* Vitamin D deficiency is the prime cause of human rickets, yet a very low intake of calcium or of phosphorus increases the susceptibility to the condition.

The disease does not develop upon a minimal mineral intake provided that the vitamin D supply is abundant, nor will a diet high in the bone-forming minerals prevent the onset of rickets or arrest its course if vitamin D deficiency exists. Rickets, for example, may develop in an infant receiving cow's milk as its sole diet or even when nursed at the breast if the mother's milk is lacking in the antirachitic vitamin.

A diet with a high proportion of cereal foods increases the susceptibility to rickets (see p 767).

(2) *Age and rate of growth* Rickets is a disease of the first two years of life, though the conditions known as *late rickets* and *osteomalacia*, which are essentially of the same nature, occur in later years (p 768). The

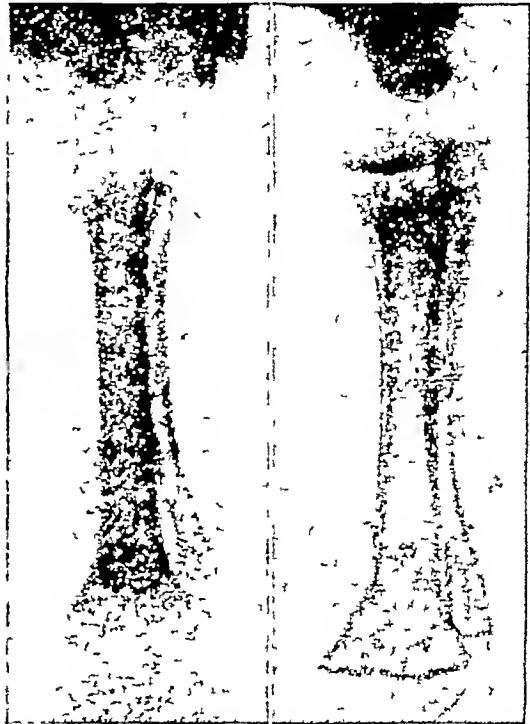


FIG 553 Left hand photograph, rickety bone, tibia of child 1½ years of age. Right, same bone after the child had received daily doses of cod-liver oil for a period of 14 days (After H. A. Harris)

disease may commence a few weeks after birth. The period during which it most frequently makes its appearance, however, is from the 3rd to the 10th month. It rarely commences after the second year, nor does it usually progress beyond this time, spontaneous healing taking place in the majority of cases around the 18th month. In some occasional instances the disease appears to commence in intra-uterine life—*fetal rickets*. *Premature infants* are highly susceptible to rickets. The possible reason for this is that since the fetus accumulates over 80 per cent of its calcium stores in the last 3 months of intra-uterine life those of the infant born before term are incomplete. Rapidly growing, overweight and apparently robust infants are more susceptible.

ble than others. Indeed Glisson (1660) to whom we owe the best classical description of the disease ascribed it primarily to over nutrition. It was noted by Hess and his associates that rats in which growth was stimulated by a more abundant diet required a greater amount of irradiation to protect them against the disease than others upon the ordinary standard diet. McCollum and associates also found that healing was induced in rachitic rats by a 5-day period of starvation. Cretins, unless stimulated to normal growth by thyroid administration do not, it is said, become rachitic. The growth impulse, therefore, appears to be essential to the development of rickets.

(3) *Climate and season.* In the days before the use of antirachitic measures had become so general, the incidence of rickets was high in the large cities of northern latitudes and for the following reasons.

Also in temperate climates it is usual for rickets to become arrested during the summer months. A seasonal fluctuation in the blood inorganic phosphorus (low in winter and high in summer) has been found in infants living in temperate zones and appears to be a general phenomenon which again points to the climatic influence upon the susceptibility to rickets. Sunlight exerts its antirachitic effect not only through direct sunshine, but also by reflection from the sky ("sky shine") and from water, light-colored buildings, etc. Animals placed out of direct sunlight but exposed to a clear bright sky are protected from rickets. (c) The atmosphere of large cities as a result of its content of smoke, dust and water vapor (humidity) is more opaque to the ultraviolet rays, and high buildings shade the streets from sunshine and sky-shine. (d) Children in temperate climates live a greater part of their time

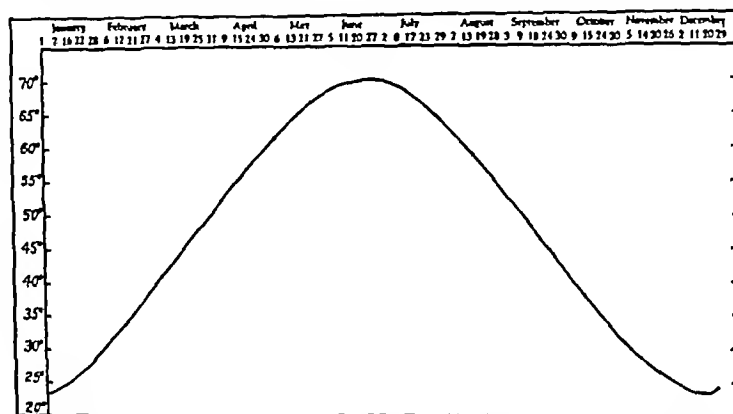


FIG 55.4 Maximum daily altitude of the sun throughout the year at the latitude of Toronto (From Tisdall and Brown)

(a) There are fewer hours of sunshine annually than in southern localities. (b) The sunshine is less intense, and the proportion of the shorter ultraviolet rays (300 to 290 $m\mu$) is smaller than in southern climates. These two factors depend upon the sun's altitude—the nearer the sun is to the horizon the greater is the depth of the atmosphere intervening between it and the earth to act as a filter. Rickets is therefore more prevalent during the winter months when the sun reaches its lowest altitude, when this is below 35° the antirachitic effect of sunlight is almost negligible, since few rays shorter than 300 $m\mu$ reach the earth. In London and Glasgow, which have been noted for their very high incidence of rickets the sun's altitude is less than 35° for 5 and 6 months of the year, respectively, whereas in such cities as Baltimore and Toronto which have shown a relatively low incidence of the disease, the altitude of the sun is less than 35° for only 3 and 4 months, respectively (fig 30c). In Jamaica and other southern localities the sun's altitude is never less than 50° and rickets is almost

indoors during the winter months and expose less of their bodies when out of doors. Ordinary window glass, it will be recalled, is opaque to rays shorter than 320 $m\mu$. Negro children living in temperate climates are especially susceptible to rickets because the skin pigment, which in their natural environment is a protection against excessive ultraviolet irradiation, further reduces the antirachitic effect of solar irradiation in northern latitudes.

Experimental rickets and the discovery of the antirachitic properties of cod-liver oil

Mellanby produced rickets in puppies by placing them upon a diet of white bread and skim milk, i.e., one deficient in the only fat-soluble vitamin known at the time, namely, vitamin A (fig 55.5). The diet included adequate amounts of vitamins B and C. He then showed that animals on such a diet, but receiving in addition cod-liver oil or

butter (which were then known to be rich sources of vitamin A), developed normally. Linseed or olive oil was not protective. From these experiments the antirachitic property of cod-liver oil was definitely established. Mellanby concluded that vitamin A, or a factor associated with it, was the antirachitic factor. Experiments by others involving the investigation of the growth-promoting and antixerophthalmic properties of vitamin A aroused the suspicion that a second vitamin was present in cod-liver oil, butter, fat, etc., and that it, and not vitamin A, was the antirachitic factor. It was found for instance that rats on a diet containing 1 per cent of cod-liver oil thrived better than those on one with 20 per cent butter-fat. A much smaller quantity of butter-fat than this

that ultraviolet irradiation had no antixerophthalmic action.

Ca P RATIOS IN THE EXPERIMENTAL PRODUCTION OF RICKETS

The experimental animal usually employed for research in rickets is the young rat (20–30 days) which unlike infants and puppies, does not, as a rule, develop rickets when deprived of vitamin D and sunlight (but some loss of bone salt may occur), if the diet is otherwise normal. If the vitamin D-free diet is also deficient in calcium or in phosphorus or the relative proportions of these are abnormal, rickets results. Unbalanced proportions of the bone forming minerals in the diet are of much greater importance than their absolute amounts. Thus a Ca P ratio which is too high or too low, though the actual amount of P or of Ca respectively is not reduced, causes the disease, which can be arrested in turn by the administration of vitamin D or by ultraviolet irradiation. In puppies, and fowl, and apparently in the human subject, an improper balance of these minerals does not seem to be an important factor in the production of rickets. Rickets is also produced in the rat by the addition of magnesium or strontium carbonate to the diet. Beryllium carbonate, likewise, when added in small quantities to a normal stock diet produces bone lesions in this species similar to those of rickets. This condition is uninfluenced, however, by vitamin D administration.



FIG. 55.5 A rickety dog. The animal was brought up on a diet deficient in the anti-rachitic vitamin D. (After E. Mellanby.)

contains sufficient vitamin A for the prevention of xerophthalmia. A factor distinct from vitamin A, therefore, appeared to be lacking in butter. It was also observed that children receiving large quantities of milk (rich in vitamin A) sometimes developed rickets. McCollum and his associates furnished the final proof of the existence of two vitamins in cod-liver oil by subjecting this material to oxidation for from 12 to 20 hours. Oil so treated had no antixerophthalmic power (i.e., no vitamin A) but was powerfully antirachitic. He proposed that this factor which was left undestroyed by oxidation should be called vitamin D. As confirmatory evidence of a separate antirachitic factor it was pointed out that cocoanut oil though containing no vitamin A furnished a certain protection against rickets. Moreover, it was shown by others

THE ANTICALCIFYING EFFECT OF CEREALS

It had long been suspected that a high carbohydrate diet was conducive to the development of infantile rickets (p. 764). Mellanby showed the aggravating effect of cereals upon the development of canine rickets. Wheat germ and oatmeal were found to be especially potent in this regard. White flour, rice and other cereals had a less pronounced anticalcifying effect. There appears to be no relationship between the rachitogenic action of these substances and their carbohydrate, protein or mineral contents. The anticalcifying effect is neutralized by vitamin D or by irradiating the cereal itself. The effect was also abolished by boiling the cereal with 1 per cent hydrochloric acid until all the carbohydrate had been converted to sugar. It is believed that Mellanby's results have an important bearing upon the development of human rickets.

Mellanby concluded that the rachitogenic action of cereals was due to their containing a toxic factor for which the term *toxamin* was suggested. The removal of the toxic effect of the cereal by boiling with HCl was thought to be due to the destruction of this hypotheti-

cal substance. The results of the later experiments of Bruce and Callow provided the basis for quite a different explanation. The anticalcifying effect is not due to a toxic substance but to the fact that the phosphorus in such a cereal as oatmeal is in an unavailable form, namely, inositolhexaphosphoric or phytic acid. Phytic acid combines with calcium and magnesium forming the complex compound phytin. Boiling HCl hydrolyzes this compound, renders the phosphorus, calcium and magnesium available and so favors calcification. Any free phytic acid present in cereals of the diet will combine with the essential bone minerals in the intestine to form the non absorbable phytin and thereby render them unavailable in nutrition. Thus, phytic acid, in a sense, is an antagonist of vitamin D (See also p 743).

Late rickets and osteomalacia

Late or juvenile rickets is essentially the same as infantile rickets but occurs in older children—4 to 16 years. It is a rare condition in western civilization but is seen in India and is then due to the same causes as those which are responsible for the occurrence of osteomalacia. *Osteomalacia* is a fault of ossification of a nature fundamentally the same as that of infantile or late rickets but occurring in adults, especially women. Pathologically, it is simply adult rickets. Nevertheless, the fact that it occurs after the period of growth and that puberty, pregnancy and lactation predispose powerfully to it give it certain features which distinguish it from the infantile or juvenile type. The entire bone is softer than the ordinary rickety bone, its total mineral content being greatly reduced. The calcium shows a greater decrease than does the phosphorus, i.e., the Ca/P ratio is reduced. The magnesium content is increased. The blood calcium is lowered and tetany in consequence is a frequent complication. The calcium balance is negative, whereas the phosphorus balance is usually normal.

The softness and pliability of the bones leads to deformity sometimes of extreme degree. Pelvic abnormalities create serious hazards during childbirth. The disease is very rare in temperate climates, though an outbreak occurred in Vienna following the first World War (hunger osteomalacia). Osteomalacia is common in India where the Mohammedan and high caste Hindu women follow the custom of purdah which demands that they live secluded within doors, they are thus deprived of the calcifying power of the sun's rays. The diet too is poor in meat, milk and vitamin D, but rich in cereals. The disease is also common in certain districts of China and is due to similar causes—a cereal diet combined with an indoor life. Cod-liver

oil or some other source of vitamin D is specific for the disease.

THE RELATION OF VITAMIN D TO DENTAL DISEASE

M. Mellanby, some years ago, carried out a series of experiments in which the effect of vitamin D and other factors upon tooth structures were thoroughly investigated. Young puppies were fed upon vitamin D deficient diets for periods extending over several months. At the end of the experimental period the animals' teeth were carefully examined with regard to gross appearance and histological structure. The following defects were observed:

- (1) Delay in the eruption of the permanent teeth
- (2) Thinning of the bony tissue of the jaws and irregularity in the arrangement of the teeth
- (3) Poorly calcified enamel which showed pitting, grooving and pigmentation

A diet of cereals—oatmeal, maize, white flour, rye or barley, especially the first of these, increased the severity of the defects. The anti-calcifying effect of cereals could be completely prevented by liberal allowances of cod liver oil or irradiated ergosterol. Irradiation of the animals with the mercury vapor lamp exerted a less pronounced beneficial influence upon the tooth structure.

It was also shown that the teeth of the offspring were influenced by the diet of the mother during pregnancy. The deciduous teeth of puppies whose mothers during pregnancy and lactation had been fed diets deficient in vitamin D erupted late and were poorly calcified. The maternal influence was also seen in the permanent teeth, the latter being much less resistant to the ill effects of a deficient diet after weaning, if the puppies had been born of a mother which received a deficient diet during pregnancy and while she was suckling her pups.

Vitamin D was shown to have an important influence upon tooth repair in adult dogs. It is not possible to induce dental defects in full grown dogs by dietary means. When, however, a dog's tooth is filed at intervals of a few days it reacts to the injury by the formation of so-called secondary dentine. The amount and quantity of the new-formed dentine were found to be very favorably influenced by vitamin D administration, cereals were, on the other hand, detrimental. In none of the experiments upon puppies or adult dogs was actual caries produced. Yet, the effects upon tooth structure and repair resulting from vitamin D deficiency are undoubtedly pertinent to the question of the development of caries in man, since such defects would presumably prepare the way for bacterial invasion.

Direct evidence that the results of deficient diets are applicable to the question of human caries has been obtained by M. Mellanby. The investigations were carried out upon English school children (6-12 years)

The children were divided into groups. All received the same basal diet but the quantities of cereal and vitamin D were varied in the different groups. The teeth were examined and their state recorded at the beginning of the experimental period and again after the lapse of several months.

Taking the experimental results as a whole, the factors influencing the development of caries, in so far as this is the result of abnormalities of tooth structure, may be summarized as follows, (a) the level of vitamin D intake or the degree of ultraviolet radiation (b) the proportion of cereal (anticalcifying) and (c) the supply of calcium and phosphorus, though this factor would appear to be of secondary importance.

It should be mentioned that the results of other investigations into the action of vitamin D in preventing dental caries are not altogether in accord with those of the Mellanby experiments (see Toverud). Though opinion is divided with respect to the value of vitamin D supplements in preventing or arresting dental caries in mature teeth, it is generally agreed that the administration of the vitamin to young children favors the development and maturation of normal tooth structure and, therefore, raises the resistance of the teeth to caries in later life.

Toxic effects of vitamin D, hypervitaminosis D. Calciferol (vitamin D₂), and to a somewhat less extent vitamin D₃, given in large doses causes rarefaction of bone, hypercalcemia, increased excretion of calcium and calcification of soft tissue. In dogs, the symptoms, though longer in making their appearance, are identical with those caused by overdosage with parathyroid extract. The first signs of toxicity are anorexia, muscular hypotonia, and extreme lassitude. These effects are followed by great prostration, collapse and death. Puppies given a *single* large dose become ill within a few days, show emaciation, refuse food, and die within two or three weeks (Taylor and Weld, Hendricks and associates). Rats, rabbits, mice and fowl are relatively immune to vitamin D overdosage. Massive doses of the antirachitic vitamin—from 300,000 to 1,000,000 international units—have been recommended in a number of conditions: pulmonary tuberculosis, lupus, postoperative tetany, psoriasis and chronic arthritis. The administration of such huge doses is not, however, free from danger. In adults, doses of over 150,000 international units, and even smaller doses per kilo of body weight in children, may be followed by alarming symptoms. There have been a number of reports of toxic effects similar to those seen in dogs: Nausea, vomiting, apathy, stupor, impaired renal function, hypercalcemia and calcification of soft tissues, e.g., blood vessels, stomach and lungs, have been described. A few deaths have occurred.

VITAMIN E (ANTI-STERILITY)

Mattill and Conklin observed some years ago that rats reared upon a diet of whole milk were

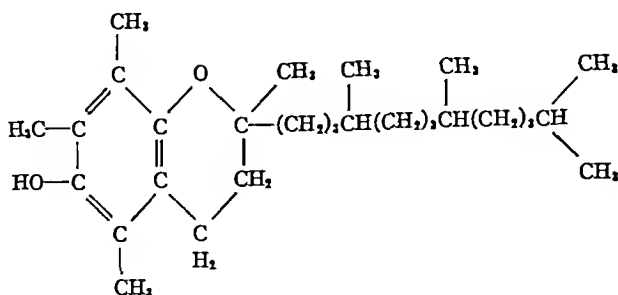
usually sterile. Evans and Bishop observed a failure of reproduction in rats (male or female) fed upon a diet consisting of casein, starch and lard, and containing butter fat or cod-liver oil as well as the other known vitamins and essential minerals. The addition to the diet of lettuce, wheat germ, or alfalfa corrected the defect. It was therefore concluded that a hitherto unknown vitamin existed which was necessary for normal reproduction. This was termed vitamin E.

Chemical properties and sources

Vitamin E is soluble in fat and in the usual fat solvents. As in the case of the other fat-soluble vitamins D and K and carotene, bile in the intestine is necessary for its absorption. Its chief sources are green vegetables, e.g., lettuce, peas, alfalfa and the germ of various seeds. Animal tissues contain little of the vitamin. Wheat germ oil has a very high vitamin E potency and most of the ordinary vegetable oils contain it in fairly large amounts. Like vitamins A and D it is contained in the unsaponifiable fraction of oils and fats. It is absent from the endosperm of wheat (white flour).

Vitamin E activity is not confined to a single compound, but like vitamins A, D and K is multiple in nature, the different compounds with vitamin E activity being known as *tocopherols* (tokos = childbirth, phero = I bear). Three physiologically active, crystalline isomeric compounds—*alpha*, *beta* and *gamma tocopherols*—have been obtained by Evans and his associates from vitamin E concentrates of wheat germ oil. α -Tocopherol, which alone is present in lettuce oil, has much the highest potency of the three. The tocopherols are associated in nature with antioxidative agents, i.e., substances which inhibit the oxidation and consequent rancidity of certain oils and fats. This antioxidative property of substances rich in vitamin E does not run parallel with their vitamin activity and is not due to the same group in the tocopherol molecule. Vitamins E are very susceptible to oxidation and ultraviolet irradiation, and rapidly lose their activity in the presence of fats and oils undergoing oxidation, a fact which is made use of in preparing E deficient diets, an oil or fat such as cod-liver oil or lard on the verge of rancidity being added in order to destroy the last trace of vitamin E activity. From their antioxidative properties and the similarity of their absorption spectra it was suspected that tocopherols and hydroquinone were closely related chemically. This suspicion was

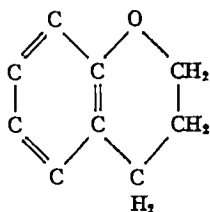
substantiated by the identification of Fernholtz of durohydroquinone among the decomposition products of α -tocopherol when subjected to heat. α -tocopherol was later synthesized by Karrer and his colleagues, and by Smith, Ungnade and Pritchard from trimethylhydroquinone and phytylbromide. The following structural formula has been proposed for it by Fernholtz



Vitamin E, α -Tocopherol ($C_{29}H_{48}O_2$)

The effect of vitamin E in enhancing the action of vitamin A has been mentioned (p 746)

Vitamin E is a derivative of *chromane*



Chromane

The effects of vitamin E deficiency

In male rats on a vitamin E deficient diet the loss of fertility is a progressive process. In the earlier stages of the vitamin deficiency the spermatozoa lose their motility, later they fail to be produced. Finally the spermatogenic epithelium degenerates and the sex instinct fails. In females, estrus occurs normally and when fertilization occurs the implantation of the ovum is not prevented but the embryos after developing for a short time die and are resorbed. The nature of the defect in embryological development leading to fetal death has been studied by Adamstone in chick embryos of hens on vitamin E deficient diets. He found that cell proliferation in the mesoderm forms a ring in the blastoderm stage which by strangling the vessels cuts the embryo off from its blood supply. Rarefaction of the mesenchyme and failure in the development of the blood-forming tissues is the

principal abnormality seen in the embryos of vitamin E deficient rats

The evidence strongly suggests that vitamin E is implicated in some way with the metabolism of the cell nucleus. The effects following a dietary lack of this vitamin are seen most prominently in actively growing structures, i.e., in cells undergoing rapid proliferation and maturation. It has,

on this account been employed to promote growth in premature infants and has been found to stimulate healing in rats

In animals the administration of vitamin E, so it is claimed, even in the absence of any obvious deficiency, exerts an effect upon the reproductive functions. For example, wheat germ oil fed to rabbits is said to increase the size of the litters, and it has been asserted that cows previously sterile have passed through normal pregnancies after receiving large doses of vitamin E. Dystrophy of skeletal muscle has been demonstrated in a number of animal species to result from vitamin E deficiency. The creatine content of the muscles is reduced and excretion of creatine in the urine is increased. These muscular defects are corrected by the administration of vitamin E. But the use of the vitamin in the muscular dystrophies of man have not been followed by much benefit. It is said to be more effective if combined with inositol. But, there is no proof that any clinical condition is due to a deficiency of this vitamin or can be cured by its administration, nor is there evidence that vitamin E deficiency occurs in man. Though earlier reports seemed to indicate that vitamin E in the form of wheat germ oil was of value in the treatment of sterility in women due to repeated miscarriages (habitual abortion), and though some support for such an action has been given by more recent work, full confirmation has not been secured. It does appear, however, that many pregnant women suffer from a deficiency of vitamin E.

It has been shown by Evans and his associates that vitamin E, or a principle closely related to it, has also a growth-promoting action upon rats after the fourth month. The influence upon growth is not secondary to an effect of the vitamin upon the sex glands since it is evident after the latter have been removed. Other manifestations of vitamin E deficiency in rats are partial paralysis of the hind limbs, due to degenerative changes in the muscles, and loss of hair. These appear only after a prolonged period (15 months) on diets low in vitamin E.

There appears to be a relationship between the pituitary and the thyroid and vitamin E. The potency of extracts of pituitaries of female rats on vitamin E deficient diets to induce ovulation in rabbits is reduced, the power is restored by the administration of vitamin E. Degenerative changes in the pituitary, involving both acidophil and basophil cells, have been described as resulting from a deficiency of this vitamin. Several of the effects of vitamin E deficiency resemble those resulting from hypophysectomy.

VITAMINS K THE ANTIHEMORRHAGIC OR COAGULATION (GERM KOAGULATION) VITAMINS
(SEE ALSO CH 17)

Chicks, young geese and ducklings upon diets lacking in green stuff are subject to a fatal hemorrhagic disease which, as a result of the work of Dam and his associates, has been shown to be due to vitamin K deficiency. It has been established by these and other workers that the bleedings are due to a reduction in the prothrombin concentration of the blood (see p. 119). The low value of this essential factor in the clotting mechanism leads to a prolonged coagulation time. Green leaves, especially alfalfa and other clovers, spinach, cauliflower and cabbage, are rich sources of vitamin K; cereals, carrots, yeast and wheat germ contain it in min-

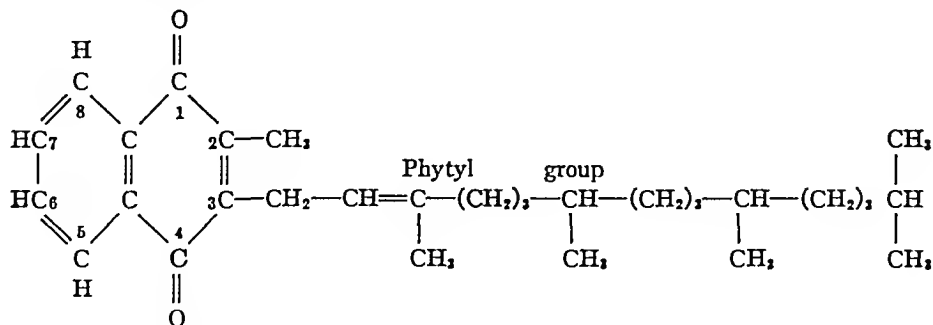
³ Inhibition of such action, as shown by hypoprothrombinemia, has been induced in rats upon diets deficient in vitamin K by the oral administration of the bacteriostatic agent, sulfaguanidine (see p. 121)

small amounts, while little or none is present in potatoes, mangels or cod-liver oil

It has not been found possible to induce a hemorrhagic tendency in any mammal so far investigated—guinea pigs, rats, hogs or dogs—by feeding K, free diets alone. No hemorrhagic disease in man, so far as is known, is due to lack of vitamin K in the diet. Almquist and Stockstad have shown that synthesis occurs to some extent in the intestine of chicks as a result of bacterial action,³ and it seems established that the immunity of man and other mammals listed above to a dietary lack of vitamin K is due either to synthesis in greater degree than occurs in the intestines of birds, or to better absorption. Bile, as shown by Greaves and Schmidt, is necessary for the absorption of the vitamin, in animals with bile fistulae and in obstructive jaundice in man, failure of the vitamin to be absorbed is the cause of the bleeding tendency associated with these conditions. The administration of vitamin K concentrates combined with bile salts is corrective (see also p. 541). Not only obstructive jaundice but any condition which interferes with the absorption or utilization of the vitamin may result in K deficiency and a tendency to bleed. Thus, in sprue and other states associated with defective fat digestion and absorption, in biliary or gastrocolic fistula or in some liver diseases, vitamin K lack may be encountered. The hypoprothrombinemia induced by dicumarol (p. 118) or by salicylic acid, and the hemorrhagic tendency of the new-born are counteracted by the administration of vitamin K.

Chemistry

McKee and his associates in 1939 isolated vitamin K from alfalfa (called K₁) and from putrefying fish meal (K₂). Later K₁ was synthesized by Binkley and by Fieser and their associates. It is 2-methyl-3-phytyl-1,4-naphthoquinone and has the following formula:



Vitamin K₁ (2-methyl-3 phytyl-1, 4-naphthoquinone) C₃₁H₄₆O₂

Vitamin K₁ is fat-soluble, it is contained in the unsaponifiable fraction of plant lipids. Vitamin K₂ (C₄₁H₅₆O₂, 2-methyl-3 (?) -1,4-naphthoquinone) is similar to K₁ in chemical structure and in its physiological activity, but has a side chain different from the phytyl group, and its potency is only about 60 per cent as great. Several other naphthoquinone compounds have been shown to possess vitamin K activity. One of these was isolated in 1933 by Anderson and Newman from tubercle bacilli and later synthesized, but it was not until 1939 that its antihemorrhagic property was demonstrated by Almquist and Klose. This substance (sometimes called phthiocol) is 2-methyl-3-hydroxy-1,4-naphthoquinone. Other naphthoquinone compounds have since been investigated. Of these 2-methyl-1,4-naphthoquinone (or *menadiolone*) is the most powerful antihemorrhagic factor known, being about three times more potent than K₁. It differs from the latter chemically only in that it does not contain the phytyl side-chain, the latter would appear, therefore, to be of no importance in so far as physiological activity is concerned. Water soluble synthetic naphthoquinone derivatives have been prepared. Some of these can be given orally without bile salts or intravenously, which of course is a decided advantage.

THE MEASUREMENT AND STANDARDIZATION OF VITAMIN POTENCIES

The general principle employed in testing a food material for its content in a given vitamin is as follows. An animal e.g., rat, guinea pig, etc., susceptible to deprivation of the vitamin to be tested is placed upon a diet which lacks that vitamin but is otherwise adequate. The substance containing the vitamin is added to this basal diet and the amount required to prevent the appearance of the effects characteristic of a lack of the vitamin under test (preventive test) or the amount required to correct the effect after it has appeared (curative test) is noted. In most cases the preventive test is employed. The minimal daily quantity of the food material required is said to contain 1 unit of the vitamin. For example, one method for testing vitamin A is as follows, a group of young (21–29 days old) healthy rats are placed upon a diet adequate in all respects (e.g., protein, carbohydrate, fat, minerals and vitamins) except that it is completely lacking in vitamin A. Another group of rats is placed upon this diet but containing, in addition, weighed amounts of the food to be tested, e.g., butter, fat, green vegetable or other material. After a short time (4 weeks or so) growth ceases in the first group of rats and the first eye symptoms appear. The minimal quantity of material

required to be fed to the second group for the prevention of these effects and to cause a gain in growth of 3 grams per week is said to contain 1 unit of vitamin A. The content of vitamin A in the various foods may therefore be expressed as the number of units per gram, per ounce or per pound, or per 100 calories. Thus, a good sample of butter contains about 22,500 units per pound and 700 units per 100 calories, whereas lettuce contains only 2400 units per pound but on account of its low calorie value, some 2775 units per 100 calories. Similar methods are employed for the measurement of the vitamins B₁, C, D and E. The prevention of arrested growth is the criterion for estimating thiamine. The prevention of scurvy in guinea pigs is the criterion for C. For vitamin E it is the power to restore fertility in a female rat in which a series of previous resorptions has occurred. In the measurement of vitamin D, young rats (40–70 grams in weight) are deprived of ultraviolet irradiation and placed for 3 or 4 weeks upon a special rickets producing diet, i.e., one with a high Ca/P ratio⁴ and lacking in the antirachitic vitamin A. A unit of vitamin D is taken as the daily amount which will prevent (preventive test) or will cure the condition after it has appeared (curative test). Evidence of the existence and extent of the disease or of the degree of healing is obtained by (a) X-ray examination, (b) analysis of the bones of their ash content or (c) by means of the "line test". Determination of the ash content or examination of the bone by means of the line test involves the sacrifice of the animal, the former method being used in the preventive test, the latter in the curative test. By X-ray examination the extent of healing under vitamin D administration can be followed in the same animal. It may be used either in a preventive or a curative test.

The "line test" depends upon the fact already alluded to on p. 764 that the end of the severely rachitic bone is mineral free. When healing commences lime salts are deposited. In the employment of this test the animals are kept upon a rachitogenic diet until severe rickets has developed. A measured dose of the material containing vitamin D is then administered daily. At the end of 10 days the animals are killed and the ends of the tibiae are split longitudinally and immersed in a 1.5 per cent solution of silver nitrate. If calcium phosphate is present in the metaphysis of the bone, silver phosphate is formed, which upon exposure to light is reduced to metallic silver, which causes a dark band to appear, the test is then designated as "positive" (+). In the absence of healing the bone end remains unchanged in color, the test is then said to be "negative".

⁴ Steenbock's rachitogenic diet is as follows

Ground yellow maize	76
Wheat gluten	20
Sodium chloride	1
Calcium carbonate	3
Its Ca/P ratio is 4 to 1	

(—) The degree of healing may be judged from the depth of the darkened band and is usually expressed by the signs +, ++ or ++++. A colorimetric method for the determination of vitamin D can be employed. The reagent is antimony trichloride and acetylchloride in chloroform. It reacts with the vitamin to give a yellowish pink color.

Riboflavin possesses the property of fluorescence. Upon this is based the fluorometric method of assay, the estimation being made upon solutions of the vitamin by a means of a fluorophotometer. An animal assay method, based upon the general principles outlined above, or a microbiological method is also available. The growth response to the vitamin of *Lactobacillus casei* is the basis of the latter method. For the assay

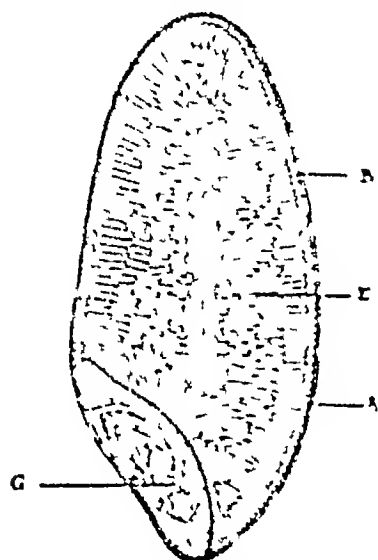


FIG. 55.6. Diagram of a longitudinal section through a grain of wheat showing B, pericarp forming the branny envelope; A, aleurone layer of cells forming the outermost layer of the endosperm removed with the pericarp during milling; C, parenchymatous cells of the endosperm; G, embryo or germ (From *Med. Res. Coun. 1 Rep.* 1932, No. 167.)

of *nicotinic acid* either a colorimetric or a microbiological method may be used. The former method is based upon the yellow color produced by the reaction of pyridine with an aromatic amine in the presence of cyanogen bromide. The microbiological method depends upon the effect of nicotinic acid upon the growth of cultures of the dysentery bacillus. A colorimetric and a microbiological method are also available for assaying *pyridoxin*. The colorimetric method is based upon the color which appears when pyridoxin is acted upon by diazotized sulphamic acid and p-nitroaniline in an alkaline medium. The microbiological method employs the fact that yeast cells or lactic acid bacilli grow only in the presence of pyridoxin. A rat growth method can also be used, in which rats are fed a basal diet complete in all respects except that it lacks pyri-

doxin. Other methods of pyridoxin assay are the curative and preventive methods based upon the development of acrodynia in rats. *Pantothenic acid* may be assayed by the chick growth test or by the growth response of certain bacteria, e.g., *Proteus morganii* or *Lactobacillus casei*.

The *folic acid* potency of foods is assayed by a microbiological method based upon its property of stimulating the growth *Lactobacillus casei*.

A reliable chemical method has been developed for the assay of vitamin C, based upon its high reducing power. The vitamin C content of a material is estimated from the reduction (decoloration) of the dye 2,6-dichlorophenolindophenol. For the detection of vitamin C deficiency clinically either the capillary resistance test (p. 122) or the intradermal test may be used. The intradermal test consists of injecting a solution of 0.0025 M dichlorophenolindophenol. The dye is decolorized in 10 minutes in subjects well supplied with vitamin C but not for 15 minutes or more in those suffering from a deficiency of vitamin C. The vitamin concentration of the plasma also may be employed, or the saturation test (p. 760).

Vitamin K may be assayed by the curative method of Almquist and Klore. This is based upon the fact that the clotting time of vitamin K deficient chicks becomes normal within three days after their being placed upon an adequate diet.

Owing to the confusion which had arisen from the employment by various workers of different standards for measuring and expressing vitamin potencies the Commission on Biological Standardization of the League of Nations recommended the following vitamin units for international usage:

- (1) Vitamin A unit = The vitamin A activity of 0.6 microgram (0.6 μ) of pure β -carotene prepared from carotene by Willstätter's method. (This is from $\frac{1}{4}$ to $\frac{1}{3}$ the amount required daily to cure xerophthalmia and restore growth in the rat.)
- (2) Vitamin B₁ unit = The antineuritic activity of 3 micrograms of thiamin hydrochloride, in colorless monochlinic plates, with a melting point, 246–247°. This is also the U. S. Pharmacopoeia unit.

No international standards have been established for the other factors of the B complex but certain units have been proposed by different laboratories. A *chick unit* has been recommended for pantothenic acid: it is 14 micrograms of pure pantothenic acid— $\frac{1}{4}$ the daily amount which will just provide maximal growth for a 3 weeks old chick on a diet free from the vitamin. A *yeast unit* based upon its requirement for the growth of yeast, and a microbiological unit based upon its effect upon the growth of *Streptobacterium plantarum* are also employed. A unit of *pyridoxine* has been defined by György as the amount of the vitamin required daily by the rat.

to cure or prevent the symptoms caused by the absence of this vitamin from the diet. A *rat unit of riboflavin* is the amount required daily by the rat to give normal growth (10 gm increase in body weight per week for from 2 to 4 weeks)

- (3) Vitamin C unit = The vitamin C activity of 0.05 mg l ascorbic acid. This is also the U.S.P. unit and is $\frac{1}{20}$ the quantity required daily to prevent scurvy in a guinea pig on a scorbutic diet
- (4) Vitamin D unit = Antirachitic activity of 0.025 microgram (0.025 γ) of crystalline vitamin D₂ (calciferol). This amount, given daily for 8 days

to a rachitic rat, causes a broad band of calcification in the metaphysis of the proximal ends of the tibiae

- (5) Vitamin E unit = The vitamin E activity of 1 mg of synthetic racemic α tocopherol acetate. This is the average amount of the latter which given orally will prevent resorption gestation in rats on a vitamin E deficient diet

There is no international standard for vitamin K. A number of vitamin K units have been proposed by different workers, e.g., Doisy, Thayer, Dam, Almquist, and others. A Doisy Thayer unit which equals 10 Dam units, is $\frac{1}{1000}$ the activity of 1 mg of pure vitamin K₁

CHAPTER 56

DIETARY REQUIREMENTS

In planning a diet the following requirements must be taken into account

- (1) The total caloric value
- (2) The proportion of the different foodstuffs—carbohydrate, fats, and protein
- (3) The mineral constituents
- (4) The vitamin content

In order that the body shall not be forced to consume its own tissues for fuel, the caloric value of the ingested food for 24 hours must balance the heat eliminated by the individual during the same period. The basis for computing the latter value is the basal metabolic rate plus an allowance for the energy expended in performing work. The basal metabolic rate is obtained by direct determination (ch 45) or by calculation from the subject's height and weight.

The average extra caloric allowances for different grades of muscular activity are

	<i>Calories</i>
Sedentary life	800-900
Light work, e g, professional and business men	900-1400
Moderate work, e g, mechanics	1400-1800
Heavy work, e g, laborers, athletes, etc	1800-4500

SAMPLE CALCULATION OF CALORIC REQUIREMENT

A young, average-sized man (surface area 1.8 square meters) has a basal metabolism of $(1.8 \times 40 =) 72$ Calories per hour, or a total of $(72 \times 16 =) 1152$ Calories for the 16 hours that he is awake. During sleep the metabolism is 90 per cent of the basal value. So $(\frac{90}{100} \times 72 \times 8 =) 518$ Calories would be his energy requirement for 8 hours of the day. Apart from that required for work his total daily requirement is therefore $(1152 + 518 =) 1670$ Calories. To this is added the appropriate work allowance given in the table above which will bring the total up from 2500 to over 5000 Calories. An allowance of 10 per cent is made to cover the increased metabolism due to the food itself (specific dynamic action, p 640). These calculations are given in tabular form below.¹

	<i>Calories</i>
Basal metabolism (16 hours)	1152
Metabolism of sleep (8 hours)	518
Allowance for light work	1200
	<u>2870</u>
10 per cent for S D A	287
Total	<u>3157</u>

The actual caloric values of various foodstuffs are given in table 75. See also table 50, for the energy requirements for different types of exercise. The effect of climate upon the metabolism must also be taken into account, a higher energy intake being required in cold and temperate climates than in the tropics.

For more precise calculation of the caloric requirements for different types of work, other factors must be taken into account, e g, the body weight and the temperature in which the work is performed. An allowance, or a deduction of 3 per cent is made, respectively, for each 10 degrees Fahr. rise or fall in the average temperature (70°F). Thus, if the total metabolism as calculated above is 3157 Cal, but the temperature is 60°F, the diet is planned for a total energy expenditure of $(3157 + 94 =) 3251$ Cal. In performing certain types of work, such as walking, ascending stairs, or any work which entails lifting the body or its heavier parts, the energy expenditure is more nearly proportional to the body weight raised to the power 0.7, than to the surface area.

THE RELATION OF AGE AND SEX TO THE CALORIC REQUIREMENT

Women have a somewhat lower basal metabolic rate (p 619) than men, and generally speaking they expend less energy in muscular work, their food requirement is proportionately less. The calorie intake should be about 20 per cent greater during the latter part of pregnancy, and about 30 per cent greater during lactation than at ordinary times.

Children require weight-for-weight a greater food allowance than the average adult for three reasons: (a) Their basal metabolic rate is considerably higher especially at the younger ages. (b) A proportion of the food material is utilized for building body tissue. (c) Children, as a rule, expend more energy in muscular activity than the average adult. For example, a boy of 16 years of age, of average physique and taking an active part in games, requires a daily energy intake equal

¹ Keys gives an approximation of the relationship between body weight and total metabolism as $\text{Calories} = AW^n$, where A is a numerical constant, W, the body weight, and n a value between 0.7 and 1.0.

to that of a man—from 3000 to 4000 Calories or from 50 to 70 Calories per kilogram of body weight per day (average adult requirement 46 Calories per kilogram) When growth is very rapid an even larger allowance may be required The food requirement of a girl of 16 to 18 years of age is approximately that of a full grown woman, though of course size, rate of growth, and the amount of exercise which she takes are modifying factors During the first two years of life the daily dietary requirement amounts to about 100 Calories per kilogram (45 Calories per pound) of body weight and from the second to the fourteenth year to about 80 Calories per kilogram

The relative caloric requirements at different ages are shown in table 75, the requirement of the average man, i e , one having an energy expenditure of 3000 Calories being taken as unity In computing the food requirements of a household this is referred to as the "man value" As a first approximation the other members of the family are apportioned Calorie allowances in accordance with the figures shown in the table Thus, the mother's allowance would be 0.83 of a "man value" or about 2500 Calories, whereas a boy of 14 years of age would receive 3000 Calories

THE PROPER PROPORTIONS OF THE THREE FOOD-STUFFS IN AN ADULT DIET

In a diet having a total energy value of about 3000 Calories the approximate amounts and percentages of its three main constituents are as follows

	GRAMS	CALORIES	PER CENT OF TOTAL CALORIES APPROXIMATE
Carbohydrate	380	1440	48
Fat	133	1200	40
Protein	70	360	12

Cereals and other carbohydrate foods contain a large proportion of water Lean meat is about 80 per cent water, whereas fats and oils are highly concentrated foods, being for the most part water free It will also be recalled that weight-for-weight fat has more than double the caloric value of pure carbohydrate or protein Over 40 per cent of the protein intake of the adult should be made up of proteins of first-class biological value

(p 644) In the diets of persons living in northern climates the calories derived from fat may constitute 45 per cent or more of the total Fat and carbohydrate are more suitable for cold climates than protein Rats, when given free choice, select carbohydrate in preference to either fat or protein when the temperature of their surroundings is lowered several degrees

The protein allowance

A great deal of controversy has centered around the question of the protein requirement of the adult Chittenden, some years ago, made a study of the subject in a series of experiments upon himself and groups of students, soldiers and athletes He showed that nitrogen equilibrium could be maintained upon a total daily intake of 25 grams or less of first-class protein² Upon an ordinary

TABLE 75
Scale of food requirement (Calories) based upon the adult male standard of 3000 Calories per day
(After Cathcart and Murray)

Man	1 00
Woman	0 83
Boy 14 up	1 00
Girl 14 up	0 83
Child 12 to 13½	0 90
Child 10 to 11½	0 80
Child 8 to 9½	0 70
Child 6 to 7½	0 60
Child 3 to 5½	0 50
Child 2 to 2½	0 40
Child 1 to 1½	0 30
Child up to 1	0 20

mixed diet containing proteins of varying biological values nitrogen equilibrium was established on an allowance of between 40 and 50 grams for a man of average weight (70 kg) This is from 0.6 to 0.7 grams of protein per kilogram Chittenden claimed that the larger protein intake of 120 grams considered by previous observers (e g , Voit) were unnecessary, if not actually deleterious to health It was contended that the renal work entailed in the excretion of large quantities of nitrogen was conducive to kidney disease Many other ills were ascribed to the excessive consumption of protein

² We have already seen that the daily urinary excretion of N on a nitrogen free diet but of adequate caloric value is around 3 grams (p 637) This represents the catabolism of 18.75 grams (3 × 6.25) of tissue protein

He also reduced the total energy intake to around 2000 Calories and stated that the more liberal diets were dictated by appetite rather than by physiological necessity. His subjects, he claimed, could carry out their usual activities just as well upon such a diet and enjoyed better health than they had previously upon their customary fare.

Since nitrogen equilibrium can be established upon a protein intake of around 40 grams per day, it *would* seem unnecessary to give more. The excess amount must obviously be catabolized to furnish energy, which can be furnished more economically by non-nitrogenous food,³ or be simply stored as carbohydrate or fat. The weak point in this argument is, as many critics of the low protein dietary have pointed out, that the ability of the body to adapt itself for a few months to a restricted protein intake is not proof that such constitutes the physiological optimum. It has also been pointed out that custom has dictated for persons in temperate climates a higher protein intake than Chittenden's standard. Pearl has estimated that the average daily protein intake per person of the population of the United States is around 120 grams. Even the diets of the Japanese and of the hardier races of India, according to McCay, contain a greater quantity of protein than that recommended by Chittenden. McCay observes that tribes in India who are accustomed to diets with the higher protein content are healthier and of better physique than those subsisting upon a more restricted protein intake. The investigation of Orr and Gilks also seems to show clearly the value of a generous protein allowance. Their report is based upon a study of two African tribes, the Masai and the Akikuyu, who though living side by side eat quite different diets. The diet of the Masai has a high protein content, it consists largely of meat, milk, and blood which they obtain from the living animal by piercing its jugular vein. The Akikuyu live upon a vegetarian diet composed of cereals (chiefly maize), legumes, plantains, sweet potatoes and other tubers, and green leaves. The members of the meat-eating tribe are some 5 inches taller and 50 pounds heavier than the vegetarians, and their muscular power is about 50 per cent greater. The Masai are comparatively free from

disease, whereas bone deformities, dental caries, anemia, pulmonary diseases and tropical ulcer are prevalent among the Akikuyu. Arthritis, however, was found to be much more common among the Masai. Though these tribes come of different original stocks there is a considerable amount of intermarriage, and the differences between them in physique and health appear to have a dietary rather than a racial basis. Similar observations have been made by Clements, who has reported that in New Guinea tropical ulcer is much less common in natives whose diet is high in protein than in those who live largely upon carbohydrate.

Such facts suggest that, in furthering the well-being of the body, protein probably plays an important part which is not revealed by short-term experiments based upon the study of nitrogen balances. There are indications that physical fitness and resistance to disease are associated with the higher protein intakes. It may be that it is only by supplying a comparatively large quantity of protein in the diet that certain vital tissues, e.g., ductless glands, can obtain adequate amounts of essential amino-acids to carry out their functions at the highest state of efficiency. There is little evidence, on the other hand, that a high consumption of protein causes renal or other diseases. Thomas records that the Greenland Eskimos are almost exclusively carnivorous, consuming enormous quantities of meat (and fat), yet renal and cardiovascular disease is not common among them. The Arctic explorers Stefanson and Anderson showed no ill-effects, either upon blood pressure or renal function, after living for twelve months upon a diet composed entirely of meat (reported by Lieb).

The effect of a high protein diet upon established renal disease is another matter. It has been shown by Smadell and Farr that the level of protein in the diet of rats has a pronounced influence upon the course of experimental nephritis. In animals upon a low (5 per cent) protein diet, the disease ran a much more favorable course than in those upon a diet containing a large proportion of protein (40 per cent). Nephritis in animals has also been reported to result from very high protein intakes, but the levels are far above anything possible in human dietaries.

It now appears that the true protein requirement for the average man lies about midway between the two extremes of Chittenden and the older observers. A generous allowance is from 70 to 100

³ Calories furnished by protein are by far the most expensive. It has been calculated that 1,000,000 Calories derived from cane sugar require for their production 0.15 acre, as compared with 17 acres necessary to produce the same amount of energy as represented by beef protein.

grams (somewhat more than 1 gram per kilogram of body weight) and about 15 per cent less than this for women. In an extensive investigation carried on over a period of six years upon some 400

sumption of approximately 77 grams of protein, i.e., approximately 1.1 gram per kilogram. In a nutrition survey in a suburb of Toronto undertaken by Ferguson and McHenry upon a large number of

TABLE 76
Average composition and energy values of edible portions of some common food materials

	PER CENT					ENERGY VALUE	
	Water	Protein N \times 6.2	Fat	Carbo- hydrate	Ash	Per kg.	Per pound
	grams	grams	grams	grams	grams	calories	calories
Meat							
Beef, round steak, medium fat	54.8	23.5	20.4		1.2	2,860	1,300
Mutton, leg roast	50.9	25.0	22.6		1.2	3,125	1,420
Pork, ham, luncheon bacon, side	49.2	22.5	21.0		5.8	2,870	1,305
	18.8	9.9	67.4		4.4	6,665	3,030
Chicken							
Broilers	74.8	21.5	2.5		1.1	1,110	505
Fish, cod, whole	82.6	16.5	0.4		1.2	715	325
Herring, whole	72.5	19.5	7.1		1.5	1,455	660
Salmon, whole	64.6	22.0	12.8		1.4	2,090	950
Trout, brook, whole	77.8	19.2	2.1		1.2	980	445
Fats							
Butter	11.0	1.0	85.0		3.0	7,930	3,605
Lard			100.0			9,285	4,220
Suet	13.7	4.7	81.8		0.3	7,790	3,540
Cheese							
American, red	28.6	29.6	38.3		3.5	4,765	2,165
Milk	87.0	3.3	4.0	5.0	0.7	715	325
Eggs, hens', boiled	73.2	13.2	12.0		0.8	1,685	765
Flour, white, wheat	11.5	11.4	1.0	75.6	0.5	3,650	1,660
Bread, white	35.6	9.3	1.2	52.7	1.2	2,650	1,205
Fruit							
Apples	84.6	0.4	0.5	14.2	0.3	640	290
Banana	75.3	1.3	0.6	22.0	0.8	1,010	460
Cherries	80.9	1.0	0.8	16.7	0.6	805	365
Grape fruit	93.6	0.6	0.1	5.7		267	120
Oranges	86.9	0.8	0.2	11.6	0.5	528	240
Vegetables							
Beans, dried	12.6	22.5	1.8	59.6	3.5	3,530	1,605
Cabbage	91.5	1.6	0.3	5.6	1.0	320	145
Lettuce	94.7	1.2	0.3	2.9	0.9	206	90
Potatoes	75.5	2.5	0.1	20.9	1.0	968	440
Sugar, granulated				100.0		4,090	1,860
Chocolate	5.9	12.9	48.7	30.3	2.2	6,295	2,860
Cocoa, powder	4.6	21.6	28.9	37.7	7.2	5,105	2,320

(Taken for the most part from Bulletin 28, Revised Edition, U. S. Department of Agriculture, 1906, by W. O. Atwater and A. P. Bryant.)

medical students living on diets of their own choice, Beard found that the average daily excretion of nitrogen in the urine was about 11 grams. This (after allowance was made for the loss of nitrogen in the feces) was calculated to represent the con-

sumption of approximately 77 grams of protein, i.e., approximately 1.1 gram per kilogram. In a nutrition survey in a suburb of Toronto undertaken by Ferguson and McHenry upon a large number of

school children, the daily consumption of protein was 80 grams for girls aged from 16 to 20 years, and over 100 grams for boys of the same age group.

Growing children require, in proportion to their weight, a considerably greater protein allowance

than do adults Up to 1 year about 16 per cent of the total Calories of an artificial diet should be furnished by protein (4 grams per kilogram of body weight)⁴ The protein requirement falls gradually until the sixth year when it amounts to about 13 per cent (2.6 grams per kilogram) of the total Calorie requirement. This value is maintained to the end of the period of growth Children also require a higher proportion of proteins of high biological value (p. 644) than do adults In young infants high-grade proteins should constitute about 100 per cent of the protein allowance (as in milk), at 1 year over 90 per cent and up to 5 years over 60 per cent From the latter age to adolescence the proportion should not be far from 50 per cent, and in adult life at least 35 per cent Women during the latter half of pregnancy and while lactating require a larger allowance

Persons undertaking heavy work, undergoing muscular training, or convalescing from wasting diseases also require a more liberal protein allowance In rigorous climates, protein food, on account of its specific dynamic action, is in higher demand than in warmer localities

The indispensability of fat

Since body fat can be derived from carbohydrate food it might be thought that dietary fat could be dispensed with On the contrary, fats quite apart from the fact that they furnish fat-soluble vitamins, are essential elements in nutrition

Fatty foods contain, besides neutral fats, and fat-soluble vitamins, certain fatty acids which are essential for health and which cannot apparently be synthesized in the body Rats placed upon a low fat diet, but containing all the known vitamins, after a time cease to grow, but growth is resumed when the essential fatty acids, even in small amounts, are added to the diet Saturated fatty acids, oleic, palmitic and stearic, which are derived from carbohydrate food, and stored in an animal's own body, have no protective action The unsaturated fatty acids of body fat are derived from the food, e.g., milk, butter and plant oils Other effects on rats resulting from a deficiency of the essential fatty acids are scaliness of the skin, caudal necrosis, emaciation, kidney lesions and early death The group of *unsaturated fatty acids* which are essential for normal growth and nutrition includes

linoleic (with 2 double bonds), *linolenic* (3 double bonds) and *arachidonic* (4 double bonds)⁵ Actually, only one of these fatty acids is essential, in the sense that it must be supplied in the diet, for, as with certain amino-acids, they can replace one another in the diet to a greater or less extent In fat-deficient animals, the blood lipids have an abnormally low iodine number In hogs on a low fat diet the linoleic acid content of the lard may be only 1.2 per cent as against a normal of 7 per cent Puppies

TABLE 77

Ash content of the edible portion of some common foods
(Modified from Lusk)

	IN 100 GRAMS FRESH SUBSTANCE						
	Iron	Calcium	Magnesium	Sodium	Potassium	Phosphorus	Chlorine
	mg	mg	mg	mg	mg	mg	mg
Beefsteak, lean	3.8	8	24	67	35	22	50
Liver (beef)	8.0	11.0					
Eggs	3.0	67	9	15	14	16	100
Milk, whole	0.2	120	11	51	142	94	120
Cornmeal	1.1						
Oatmeal	3.7	93	127	81	380	380	35
Rice, polished	0.7	8	27	21	68	89	50
Wheat flour	1.5	26	30	69	146	86	76
Wheat, entire grain	5.2	44	170	106	515	469	88
Beans, lima, dried	7.2	71	187	245	1743	336	25
Beans, string, fresh	1.6						
Cabbage	0.9	49	14	20	243	27	13
Corn, sweet	0.8						
Peas, dried	5.6	100	145	118	880	397	40
Potatoes	1.2	11	22	19	440	61	30
Spinach	3.8						
Turnips	0.6	64	169	59	332	51	40
Apples	0.3	10	8	15	125	13	4
Raisins	3.6	57	9	141	830	126	70

upon a diet low in fat develop a scaly or eczematous condition of the skin that is readily cured by feeding lard or other fats The serum lipids show an iodine number about 25 per cent below normal It is thought that certain eczematous conditions in the human subject are due to fatty acid de-

⁵ Unsaturated fatty acids combine with iodine in proportion to the degree of unsaturation The iodine number (number of grams of iodine which will combine with 100 grams of the fat) is, therefore, a measure of the degree of unsaturation of the fatty acids in the fat molecule.

⁴ In mother's milk the protein calories amount only to 11 per cent of the total The protein is, however, of higher quality than that of cow's milk.

iciency These subjects show a low content of unsaturated fatty acids in the blood lipids, and babies kept for several months upon a diet very

On account of their high energy value, fatty foods are demanded in relatively large amounts by men performing very heavy work, especially in

TABLE 78
Recommended dietary allowances*
(Food and Nutrition Board, National Research Council)

	CALORIES	PROTEIN GRAMS	CALCIUM GRAMS	IRON	VITAMIN A†	THIAMIN (B ₁)	RIBO- FLAVIN	NIACIN (NICO- TINIC ACID)	ASCOR- BIC ACID	VITAMIN D
				mg	I U	mg †	mg	mg	mg †	I U
Man (70 kg)										
Sedentary	2500	70	1 0	—	—	1 5	2 2	15	—	—
Moderately active	3000	70	0 8	12	5000	1 8	2 7	18	75	**
Very active	4500	70	—	—	—	2 3	3 3	23	—	—
Woman (56 kg)										
Sedentary	2100	—	—	—	—	1 2	1 8	12	—	—
Moderately active	2500	60	0 8	12	5000	1 5	2 2	15	70	**
Very active	3000	—	—	—	—	1 8	2 7	18	—	—
Pregnancy (latter half)	2500	85	1 5	15	6000	1 8	2 5	18	100	400 to 800
Lactation	3000	100	2 0	15	8000	2 3	3 0	23	150	400 to 800
Children up to 12 years										
Under 1 year‡	100/kg	3 to 4/kg	1 0	6	1500	0 4	0 6	4	30	400 to 800
1-3 years¶	1200	40	1 0	7	2000	0 6	0 9	6	35	**
4-6 years	1600	50	1 0	8	2500	0 8	1 2	8	50	—
7-9 years	2000	60	1 0	10	3500	1 0	1 5	10	60	—
10-12 years	2500	70	1 2	12	4500	1 2	1 8	12	75	—
Children over 12 years										
Girls, 13-15 years	2800	80	1 3	15	5000	1 4	2 0	14	80	**
16-20 years	2400	75	1 0	15	5000	1 2	1 8	12	80	—
Boys, 13-15 years	3200	85	1 4	15	5000	1 6	2 4	16	90	**
16-20 years	3800	100	1 4	15	6000	2 0	3 0	20	100	—

* Tentative goal toward which to aim in planning practical dietaries, can be met by a good diet of natural food. Such a diet will also provide other minerals and vitamins, the requirements for which are less well known.

† 1 mg thiamin equals 333 I U, 1 mg ascorbic acid equals 20 I U.

‡ Requirements may be less if provided as vitamin A, greater if provided chiefly as the pro vitamin carotene.

§ Needs of infants increase from month to month. The amounts given are for approximately 6-8 months. The amounts of protein and calcium needed are less if derived from human milk.

¶ Allowances are based on needs for the middle year in each group (as 2, 5, 8, etc.) and for moderate activity.

** Vitamin D is undoubtedly necessary for older children and adults. When not available from sunshine, it should be provided, probably up to the minimum amounts recommended for infants.

Further Recommendations, Adopted 1942

The requirement for *iodine* is small, probably about 0.002 to 0.004 milligram a day for each kilogram of body-weight. This amounts to about 0.15 to 0.30 milligram daily for the adult. This need is easily met by the regular use of iodized salt, its use is especially important in adolescence and pregnancy.

The requirement for *copper* for adults is in the neighborhood of 1.0 to 2.0 milligrams a day. Infants and children require approximately 0.05 per kilogram of bodyweight. The requirement for copper is approximately one tenth of that for iron.

low in fat developed a generalized eczema which was corrected when the fat of the diet was restored. During the period of low fat feeding, the iodine number of the serum lipids of these infants was depressed (See Burr and associates).

cold climates. A pound of beef fat, which contains little water, yields about 4000 Calories. The caloric value of an equal amount of white bread, which contains over 40 per cent of water, is only about one quarter of this. Fat pork, beans and peas

(vegetables relatively rich in fat) are prominent items in the diet of Canadian lumbermen and construction workers. The daily energy expenditure of some of the former workers may run to 8000 Calories. Another advantage of fat is its superior "staying power." Its digestion and absorption are extended over a much longer period than those of carbohydrates. Hunger, "emptiness" and fatigue are experienced much sooner upon a diet high in carbohydrate than upon one containing a liberal allowance of fat. The actual efficiency of fat as a fuel for muscular work as shown by Krogh and Lindhard (p. 728) and by Murlin and Marsh is only from 10 to 12 per cent less than that of carbohydrate.

Children, especially those under the age of 1 year, require a larger proportion of fat in the diet than do adults. Holt and Fales place the daily requirement at about 4 grams per kilogram up to the age of 1 year (from 35 to 40 per cent of the total Calories) with a gradual reduction to about 3 grams per kilogram at 6 years. According to these observers, a liberal allowance of fat also favors the absorption of calcium.

Mineral constituents (see also pp. 76, 814 and 859)

The minimal daily requirements of calcium, phosphorus and iron for the average adult (70 kgm. in weight) are

Calcium	1.0 gram
Phosphorus	1.5 gram
Iron	15-20 mg

In childhood and in pregnancy (especially in the later months) and in lactation the calcium requirement is higher than that given above. For infants and children up to 12 years the minimum is placed at 1.0 to 1.2 gram per day, and from 13 to 20 years at 1.4 for boys or from 1.3 to 1.0 for girls. In the later months of pregnancy, when from 20 to 30 grams of the mineral are being deposited in the fetus, daily allowance should be at least 1.5 grams,

2 grams should be the daily allowance during lactation. The best source of calcium for the growing child is milk, which contains about 1.3 grams per quart. Other dairy products, e.g., cheese, ice cream, are excellent sources of calcium. Cereals also are rich in calcium though much of it is not absorbed and retained, meat contains insignificant amounts of this element (see table 77) (p. 779). Egg yolk, meat, liver and kidney, certain vegetables, fruits and nuts are the main sources of iron. Milk is very poor in this element.

The daily requirement of iron for children is, on a weight basis, higher than for adults, milk-fed infants therefore tend to develop anemia (p. 83) after the fourth month or so unless given iron in inorganic form.

It is unnecessary, as a rule, to pay attention to the phosphorus content of the diet provided that the protein and calcium of the diet are adequate, for protein is relatively rich in phosphorus, and the latter is associated with calcium in milk, eggs, cereals, legumes and several other foods.

The basic elements, sodium, potassium and magnesium, are derived chiefly from cereals, fruits and vegetables. Sodium chloride is also taken in the form of cooking and table salt. The average daily intake is from 10 to 12 grams. Many preparations of table salt are also sources of iodine, since it has become the custom of the manufacturer to add minute quantities of this element to his product. The daily iodine requirement for adults is from 100 to 150 micrograms. Other elements (trace elements), e.g., copper, cobalt (for erythropoiesis), fluorine (for tooth and bone structure), zinc (for the action of carbonic anhydrase and insulin) and manganese (for erythropoiesis and the action of phosphatase) are essential, but are present naturally in the diet in adequate amounts.

The dietary allowances recommended by the food and nutrition board of the National Research Council, Washington, are given in table 78.

SECTION VII THE DUCTLESS GLANDS OR ENDOCRINES

(THE ENDOCRINE FUNCTION OF THE PANCREAS IS DEALT WITH IN CHAPTER 49)

By N B T

CHAPTER 57

INTRODUCTION THE PITUITARY BODY

DEFINITIONS NATURE OF HORMONES METHODS OF INVESTIGATING ENDOCRINE FUNCTION

The products of the ductless glands—thyroids, adrenals, pituitary, etc.—belong to a class of physiologically active chemical substances known as *internal secretions or hormones*. The latter term was first used by Bayliss and Starling in the report of their discovery of secretin (p 530), it is derived from a Greek word meaning to excite. Though nearly all the actions of the secretions of the ductless glands are excitatory in character, a few internal secretions (e.g., enterogastrone, p 512) are inhibitory. The term *chalone* has been suggested for the latter by Schafer but is not often used. A hormone may be defined as a chemical substance which, having been formed in one part of the body, is carried in the blood stream to another organ (the target organ) or tissue and influences its growth, nutrition and function. With the exception of the secretions of the thyroid, gonads, and adrenal cortex, the various hormone preparations, which so far have been obtained by extraction and employed in medicine, are almost or quite inert when administered orally. Even though the sex hormones and the principles of the adrenal cortex are effective by oral administration, the dose when given in this way must be very much larger than if given parenterally in order to produce the same response. Adrenaline, insulin and the principles of the parathyroids and pituitary, are effective only when given by injection. Most of the ductless glands are present in all orders of vertebrates and an extract obtained from the gland of one order exerts, as a rule, its specific effect when administered to a member of another order, e.g., the hormone of the sheep's thyroid influences the growth and development of frog larvae (tadpoles) (see also p 787). Hormones have also been demonstrated in insects, crustaceans, and even in plants (auxins and phytohormones).

Apart from small amounts which may be held in the endocrine organs themselves, hormones are not

stored in the body. Therefore, in cases of endocrine deficiency, repeated small doses rather than large doses at infrequent intervals are required to correct the deficiency. A hormone does not stimulate the gland which secretes it. Thyroid extract, for example, does not stimulate the thyroid, and the ovarian hormones do not stimulate the ovary directly, the most probable effect will be inhibition of the activity of the gland and its ultimate atrophy.

Three main methods are available for the study of an endocrine function. First, an extract of the glandular tissue may be prepared which may then be injected into animals and a study made of its effects. Secondly, a particular gland may be removed from an animal and the subsequent life history of the animal watched, careful notice being taken of its subsequent development, growth, or any unusual symptom. Or, thirdly, studies may be made upon human subjects in whom some gland is known to be deficient or overactive.

The fundamental rôle which a hormone plays in the tissue upon which it acts, whether serving as part of an enzyme system, as has been demonstrated for most vitamins, or whether it furnishes some essential metabolite, has not, except in the case of the thyrotrophic hormone (p 807), been defined.

A word should be said here concerning the general chemical nature of the active principles of the endocrines. The hormones of the sex glands and adrenal cortex are steroids and closely allied in chemical structure, while the active principles of the pituitary, thyroid and parathyroids are proteins.

THE PITUITARY BODY OR HYPOPHYSIS CEREBRI

The average weight of the human hypophysis is about 0.5 gram. Its average dimensions are $10 \times 13 \times 6$ mm, it is larger in females, especially in those who have borne children, than in males. It depends from a short stalk from that part of the base of the brain known as the tuber cinereum, and which forms in

part the floor of the third ventricle. The optic chiasma lies in front of the tuber cinereum, the mammillary bodies are behind. The pituitary body consists of a posterior and an anterior part readily separable along a natural line of cleavage which represents the fusion of the two parts during development, as described below. It is ensconced within the small cavity formed by the *sella turcica* of the sphenoid bone which is closed above by a thin membrane—*diaphragma sellae*—except where it is pierced by the pituitary stalk.

Until recent years the two parts of the pituitary body were generally referred to as the *anterior*, and *posterior lobes*, and the stalk as the *infundibulum*. Though these terms are still used somewhat loosely, it is preferred, for more precise description of the hypophysis, to employ a nomenclature based upon the origins of its different parts.

The pituitary body has a dual origin—from the ectoderm in the roof of the primitive mouth (stomodaecum) just in front of the buccopharyngeal membrane, as well as from the base of the brain. In the early embryo a pouch grows upwards from the stomodaecum to meet a corresponding hollow diverticulum prolonged downwards from the floor of the third ventricle. The evagination from the primitive mouth is called the *craniopharyngeal canal* or *pouch of Rathke*. Through pressure upon its posterior wall by the down-growth of nervous tissue, this wall of Rathke's pouch is approximated to the anterior wall and the cavity almost obliterated, a narrow cleft alone remaining in most adult animals. In man, a row of vesicles rather than a unilocular cleft is found, while the original cavity of the neural element completely disappears, this part including the stalk is solid, but in some adult animals, e.g., the cat, the cavity in this part of the pituitary persists and communicates with the third ventricle.

The part of the pituitary which originates from Rathke's pouch has for the most part a glandular structure and is therefore called the *adenohypophysis* (anterior lobe). The portion developed from the brain and fused with the adenohypophysis is known as the *neurohypophysis*. Its two parts, the stalk or infundibulum, and the expanded distal portion (posterior lobe), are now referred to, respectively, as the *neural stalk* and the *neural lobe*. The extent to which the adenohypophysis makes contact with the neural lobe varies considerably. It may almost entirely surround the latter. But in all instances it invests the stalk with a thin layer of tissue which extends to the base of the brain to cover the median eminence of the tuber cinereum. This extension of the adenohypophysis is called the *pars tuberalis*. The thin strip of tissue lining the posterior wall of the cleft and fused with the neural lobe, and which is the remnant of the posterior wall of Rathke's pouch, is known as the *pars intermedia*. The *pars intermedia* owing to its intimate relation-

ship with the neural element was usually included as part of the posterior lobe when this term was generally used, but obviously, when its origin and structure are borne in mind, it should be considered part of the *adenohypophysis*.

The remaining, and much larger, part of the adenohypophysis is called the *pars distalis*, it is also sometimes known as the *pars anterior* or *pars glandularis* (fig 57 1).

Origin	Primary Division	Sub-Divisions	
From primitive buccal cavity	Adenohypophysis	Pars distalis (Pars anterior or glandularis)	Anterior lobe
		Pars tuberalis	
		Pars intermedia	
From floor of third ventricle	Neurohypophysis	Neural lobe (pars nervosa)	Posterior lobe
		Neural stalk	

The *pars distalis* is richly vascular, showing numerous blood sinuses between cords of cells. The cells fall into two main groups, (1) *chromophobe* or *reserve* cells which possess no granules, stain lightly and apparently do not secrete, (2) *chromophil* cells which contain large numbers of granules which stain readily, they are believed to elaborate the secretion of the anterior lobe. On a basis of the character of the granules, the chromophil cells are grouped again into two varieties, (a) *acidophil* (or *alpha*) cells which stain more readily with acid dyes, and (b) *basophil* (or *beta*) cells which have a greater affinity for basic stains. The three types of cell (chromophobe and the two types of chromophil cell) are scattered indiscriminately throughout the *pars distalis*. The proportions found by Rasmussen in man, were around 50 per cent chromophobes, 35 per cent acidophils and 15 per cent basophils. The last are increased after castration (p 887).

It now appears that the chromophobes represent an earlier stage in the development of the granular (chromophil) cells and give rise to either acidophils or basophils. The *pars intermedia* of man often shows cysts of various sizes containing a hyaline or colloid material, and a few cells filled with the same material (*hyaline bodies*). The *pars tuberalis* resembles the *pars anterior* in being constituted of cords of cells separated by blood sinuses. The cells, however, are non-granular.

The *neural lobe* and *stalk* are composed of (a) neuroglial cells, (b) fusiform cells with several processes and containing granules of a brown pigment in their cytoplasm (*pituicytes*), (c) numerous nerve fibers, and (d) hyaline bodies. The hyaline bodies were believed by Herring, Cushing and others to represent cells of the *pars intermedia* and *pars tuberalis* which have undergone a hyaline change and are traversing the *pars nervosa* and stalk to reach the 3rd ventricle. The

hyaline material with which they are filled has been thought to represent the hormone of the neural (posterior) lobe. Some doubt has been thrown upon this view and many consider these bodies simply as artefacts.

Wislocki and King have investigated the blood supply of the hypophysis of the monkey and of man. The following description is based upon their observations. Except for a few capillary anastomoses, the blood supply of the pituitary is independent of that of the brain proper. It is derived from the hypophyseal arteries, branches of the internal carotid. The circulation of the pars distalis and pituitary stalk is also separate from that of the neural lobe. The stalk receives its blood supply from the superior hypophyseal arteries which form a rich plexus surrounding it and

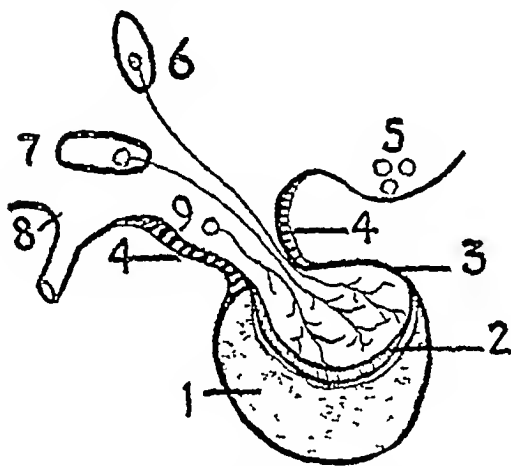


FIG 571 Diagram of the pituitary and neighboring structures 1, pars distalis, 2, pars intermedia, 3, neural lobe, 4, pars tuberalis covering neural stalk, 5, mammillary body, 6, paraventricular nucleus, 7, supra-optic nucleus, 8, optic chiasma, 9, tuber cinereum. Nerve fibers are constituents of the hypothalamo-hypophyseal tract.

lying between it and the thin overlying layer of the pars tuberalis. This vascular plexus sends twigs into the substance of the stalk and extends upwards into the median eminence of the tuber cinereum, where it anastomoses with the general capillary bed of the hypothalamus. The plexus is drained by vessels (portal venules) which descend the stalk and empty into the sinusoids of the pars distalis. The sinusoids are also supplied with blood through the superior hypophyseal arteries. Thus, as Popa and Fielding pointed out some years ago, the pars distalis receives a double vascular supply, a *portal*, the blood passing through two capillary beds, and a *systemic*. The former has been named by Wislocki the *hypothalamo-hypophyseal portal system*.

THE NERVE FIBER CONNECTIONS OF THE HY-

POPHYSIS. The *neurohypophysis* is supplied with fibers derived from the internal carotid nerve plexus. The cell stations of these fibers are probably in the superior cervical ganglion of the sympathetic, they reach the neurohypophysis along blood vessels which descend the neural stalk. Parasympathetic innervation of the hypophysis from the great superficial petrosal nerves which pass through the internal carotid plexus has also been described. The most important nervous connections of the neurohypophysis are fibers originating in the hypothalamic nuclei—supra-optic, tuber cinereum and, probably, the paraventricular nuclei (p 1024). These fibers, which according to one estimate number about 50,000, descend in the neural stalk for distribution to the neural lobe. A few can be traced to the pars intermedia but none apparently enters the pars distalis. This tract of fibers, of great importance in respect to the functions of the neural lobe, is known as the *hypothalamo-hypophyseal tract*. A few nerve fibers have been described penetrating a short distance into the pars distalis, but they are probably vasomotor in function. The secretory function of *adenohypophysis*, unlike that of the neural lobe, is probably not under direct nervous control, i.e., by secretory nerves (see p 791 and fig 571).

THE PHYSIOLOGY OF THE PITUITARY THE ADENOHYPHYSIS

The adenohypophysis is the master gland of the endocrine system. At least eleven types of physiological effect have been identified with this part of the pituitary. Six of these, since they can each be produced by an extract relatively free from other effects, are generally attributed to distinct and separate hormones, namely, (1) the *growth hormone*, (2) the *thyrotropic* (or *thyrotropic*)¹ hormone, (3) the *adrenocorticotrophic* (or *adrenocorticotrophic*) hormone, (4) and (5) the *gonadotrophic* (or *gonadotrophic*) hormones, and (6) *prolactin* or the *lactogenic hormone*².

¹ The Third International Conference on the Standardization of Hormones has recommended that the suffix *-tropic* in the adjective qualifying the hormones of the anterior pituitary be replaced by *-trophin* (see Collip), and that the termination *-trophin* be used in forming the name of the hormone (e.g., thyrotrophin). The suffix *-tropic* is from the Greek *τρέφειν*, to turn, and in this sense has been used in such terms as heliotropic, geotropic etc. The suffix *-trophin* is derived from the Greek word *τροφήν*, to nourish or nurture and is, therefore, more appropriate as an ending for those pituitary hormones which affect the development and growth of other endocrines. The *-trophin* ending is also accepted usage.

² Though it is convenient and customary to speak of these pituitary principles as separate hormones there is no means of knowing whether the active materials obtained by rather drastic methods of extraction are

The remaining effects of anterior lobe extracts, but for which separate hormones have not been definitely demonstrated, are (7) *ketogenic*, (8) *insulin antagonizing or glycotropic* (9) *diabetogenic*, (10) *parathyrotrophic* and (11) *pancreatrophic*. These various actions of anterior pituitary extracts can be divided into two main groups (a) those such as the effects upon growth, milk secretion, fat metabolism, etc., which are exerted upon peripheral tissues, and (b) those which act upon other endocrines, e.g., thyroid, adrenal cortex, gonads, etc., these latter pituitary principles are given names composed in each instance of the name of the gland acted upon and the suffix *trophic* or *tropic*.

Not a great deal is known of the chemistry of the anterior lobe principles. They are protein in nature, and it appears that the gonadotrophic hormones are glycoproteins containing glycosamine and a hexose. A polypeptid with high adrenocorticotrophic potency has been obtained (p 788).

The effects of hypophysectomy are for the most part attributable to the withdrawal of the actions of the anterior lobe secretion upon various physiological functions. In competent hands complete ablation of the pituitary is not the fatal operation it was once believed to be, animals survive for long periods during which they can be studied, though undoubtedly their life span is shortened as a result of the abnormalities which result. Among the chief effects of hypophysectomy are (1) arrested growth, (2) atrophy of the gonads and, indirectly, of the accessory organs of sex, (3) suppression of milk secretion and involution of the mammary glands, (4) atrophy of the thyroid and adrenal cortex and probably degenerative changes in the parathyroids, (5) lowered metabolic rate, (6) hypoglycemia and increased sensitivity to insulin, reduction in liver and muscle glycogen, (7) depression of spontaneous activity, (8) diminished resistance to infections and shock.

actually secreted as separate entities by the living gland. Only three characteristic types of cell are known to exist in the anterior lobe, and only two—acidophil and basophil—are believed to elaborate physiologically active principles. It seems scarcely credible that the several extraction products of the anterior lobe represent a corresponding number of separate hormones, and are not, as Riddle suggests, "hormone fragments", or as Collip expresses it, "chemically dissected" parts which represent prosthetic groups split off from perhaps only two large hormone molecules.

The growth hormone

It has been known for many years as a result of the study of growth abnormalities in man (acromegaly, gigantism and dwarfism) that the anterior lobe of the hypophysis influences skeletal growth. Experimental work also indicated such a function, removal of the posterior lobe and part of the anterior from puppies resulting in dwarfing, sexual infantilism and obesity (Aschner, Cushing). Removal of the anterior lobe in tadpoles was shown by Smith and by Allen to result in retarded growth, failure of metamorphosis and a reduction in pigmentation of the body surface.

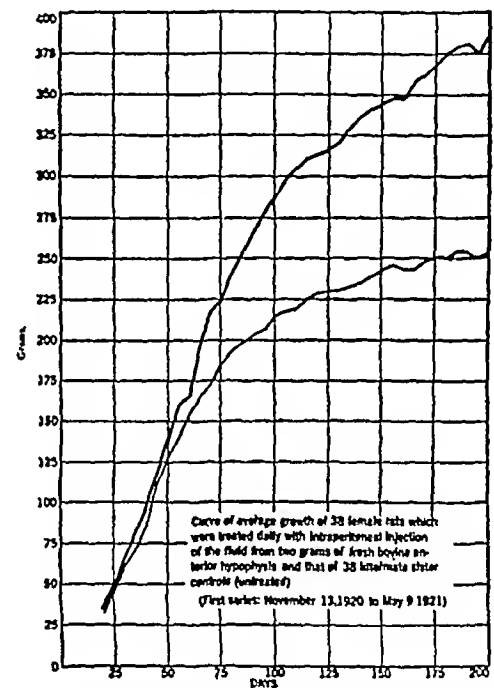


FIG 57.2 Curve of average growth of 38 female rats which were treated daily with intraperitoneal injections of anterior lobe extract, and that of 38 littermate sister controls (untreated) (After Evans)

The final proof that the anterior lobe secreted a growth-promoting hormone was furnished by the experiments of Evans and Long, who found that a saline or alkaline extract of the gland when given daily to rats by intraperitoneal injection (but not by mouth) extended the growth period beyond the normal limit. Some of these animals finally attained a size larger by 100 per cent or more than that of their untreated litter mates (see fig 57.2). The effect was usually greater in females. Similar results have been obtained in dogs. Gigantism, acromegaly and splanchnomegaly (p 799) have been produced in bulldogs by Put-

nam, Benedict and Teel. Airedales, Boston terriers and dachshunds have been shown by others to respond to such an extract (see fig 57.3) Potent extracts free from gonadotropic and containing only traces of thyrotrophic, adrenotrophic and lactogenic principles have been prepared. A specific stimulating effect of the growth hormone upon epiphyseal cartilage of hypophysectomized rats has been demonstrated. The cartilage cells become enlarged, active osteoblasts appear, and a zone of newly formed bony trabeculae develops in the previously quiescent bone. As might be expected, the stimulus to growth is accompanied by a retention of nitrogen (positive nitrogen balance). An increase in reserve or deposit protein, especially in the hypertrophied muscles, appears to be one of the first changes in protein metabolism caused by the growth principle. The greater muscle bulk appears to be due to an increase in non-contractile material for neither the myosin content nor the

whether these various metabolic effects are essentially and specifically related to the growth hormone or are due to separate and distinct principles.

The effect of hypophysectomy upon the growth of puppies is shown in fig 57.4. The potency of a given anterior lobe extract in growth-promoting properties may be assayed by injecting it in measured amounts into rats whose growth has been completely arrested by subjecting them to hypophysectomy at the age of 100 days and thereafter following their growth curves. Normal rats are much less reactive to the hormone than the operated animals.

The pituitary of a full-grown animal contains as much growth hormone as those of young animals, but the rôle which it plays in the adult body is unknown. It may be mentioned, however, as an isolated observation, that under its influence the growth of hair over a shaved part is more rapid than usual.



FIG 57.3 Litter-mate female dachshunds 11½ months old. HY-26 injected with growth hormone for 35 weeks. HY-27 control. (After Evans and associates.)

contractile power of the muscle is increased. There is no appreciable change in the calcium balance. The growth principle is almost certainly elaborated by the acidophil cells. The pituitaries of dwarf mice are lacking in acidophils while acromegaly and gigantism are frequently associated with tumors composed of these elements.

Diabetogenic and ketogenic effects are associated with the growth hormone and cannot be completely separated from it. Diabetogenic action can be demonstrated even in the most highly purified preparations, and is apparently, therefore, a property of the growth hormone molecule itself, yet this action is probably of little physiological importance since the hormone in appropriate dosage stimulates growth in young animals without inducing diabetes (young). The growth hormone exhibits other metabolic effects, namely, depression of the respiratory quotient, a decrease in arginase of the liver and of insulin in the pancreas. Also, the specific dynamic action of protein appears to be dependent upon this principle. It is not known



FIG 57.4 Hypophysectomized puppy (right) and litter-mate control (left) three months after removal of the hypophysis when three weeks old. (After Dandy and Rickert.)

It might with reason be asked, "Does the growth principle act indirectly by stimulating other glands which are known to be influenced by the pituitary, e.g., the thyroid, adrenals or gonads?" Such an indirect mode of action is disproved by the fact that growth cannot be induced in an hypophysectomized animal by the administration of an extract of any of these glands. On the other hand, growth can be induced in hypophysectomized animals or stimulated

in young normal animals by means of pituitary extracts containing only traces of "trophic" principles. The growth hormone should not, however, be looked upon as the supreme factor controlling growth in the sense that cell multiplication and the growth of tissues and organs are arrested when it is withdrawn. That the growth of individual organs is not abolished is shown by the facts that in the absence of the pituitary, mitotic figures appear in the mammary glands of parturient

e.g., increase in heart rate, rise in metabolic rate, increased susceptibility to oxygen want and increased tolerance to acetonitrile have been observed by various workers. This pituitary hormone prevents the thyroid atrophy which follows hypophysectomy. It decreases the iodine content of the gland in normal animals while raising that of the blood. The latter effect is the result, apparently,

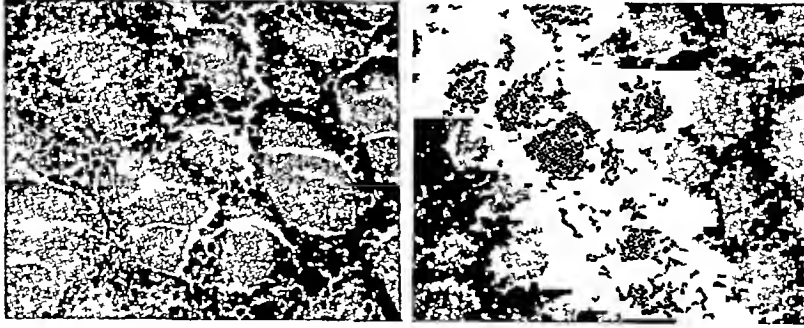


FIG 57.5 The thyrotrophic hormone. Section of thyroid from normal guinea-pig on the left, that of litter mate treated with 8 mg of anterior lobe substance on right (After Van Dyke, *The physiology and pharmacology of the pituitary body*, Chicago University Press)

rats, and, hypertrophy of the remaining kidney follows unilateral nephrectomy. The pituitary appears to preside over the growth of the body as a whole and to control the proportional growth of the several organs and parts.

The thyrotrophic hormone, thyrotrophin, thyroid stimulating hormone (TSH)

Hypophysectomy causes atrophy of the thyroid. On the other hand, an acid extract of the anterior lobe of the pituitary has been prepared (Loeb and Bassett, Aron) which, when injected into animals, stimulates the thyroid. A homogeneous substance with a high thyrotrophic activity, and which appears to be a pure protein (a pseudoglobulin), has been isolated more recently from beef pituitary tissue. It is free from growth and gonad-stimulating properties. Striking changes in the histological appearance of the thyroid result from the injection of thyrotrophin, the colloid material disappears, the epithelium becomes hyperplastic and the alveolar cavities are much reduced in size or almost obliterated as a result of the increased height of the epithelium and the infolding of their walls (fig 57.5). Mitotic figures (especially after the administration of colchicine (p 897) appear in the thyroid in large numbers, whereas in the normal gland of the guinea-pig about 150 mitoses can be found, nearly 200,000 may appear after the administration of a potent thyrotrophic principle. Signs of hyperthyroidism,

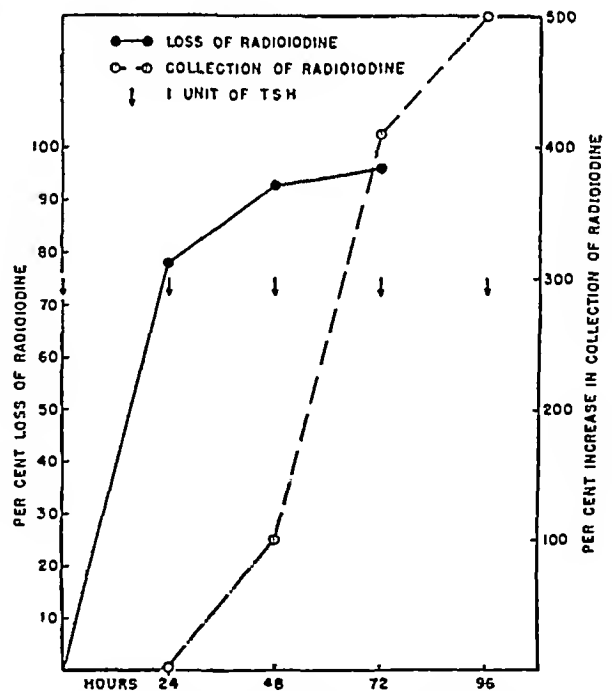


FIG 57.6 Showing the percentile loss of radio-iodine from and the delayed increase in radio-iodine in the thyroid under the influence of TSH. (After Rawson)

of the release of thyroid hormone into the circulation. When iodine is administered in pharmacological dosage, its uptake by the gland is increased by TSH, but this effect is delayed for about 48 hours after the injection. The production of thy-

roid hormone is also increased at this time (see also ch 58)

The thyrotrophic hormone induces creatinuria and increased calcium excretion, and a reduction of liver glycogen in ordinary laboratory animals, and hastens the metamorphosis of tadpoles—all well-known thyroid effects. None of the effects so far listed can be produced after removal of the thyroid. The action of thyrotrophin has been demonstrated in man, injections causing a rise in the metabolic rate in normal persons, hyperthyroid patients and those suffering from pituitary deficiency, but not in subjects with functionless thyroids. *In vitro*, this hormone causes hyperplasia of sliced thyroid tissue, evidence that its action is

came refractory. The responsive period was followed by one during which the B M R fell to normal or to a subnormal level (though the injections were continued), the thyroid hyperplasia, however, persisted. Eventually the thyroid returned to its normal size and histological appearance, in spite of the uninterrupted administration of the extract. Slight prominence of the eyeballs appeared toward the end of the responsive phase, but the most pronounced exophthalmos occurred during the refractory phase. The production of the effect upon the eyes appears, therefore, to be independent of the effect upon the B M R. It has been mentioned elsewhere (chapter 58) that exophthalmos has been induced by the administration of anterior pituitary extract to thyroidectomized animals.

The discovery of the thyrotrophic hormone and the observations just cited suggest at once the possibility that exophthalmic goiter and certain instances of hypothyroidism are primarily of pituitary origin—excess and deficiency, respectively, of the thyrotrophic principle.

The adrenocorticotrophic hormone (ACTH), corticotrophin

Hypophysectomy in the rat leads to atrophy of the cortex of the adrenal (fig 577). Hypertrophy of the adrenals, chiefly the cortex, is observed in hyperactivity of the anterior pituitary in man (see acromegaly and pituitary basophilism, p 799), whereas pituitary cachexia (p 801) is associated with adrenal atrophy. Adenomatous tumors of the adrenal cortex have been produced by the injection of anterior pituitary extracts into normal animals. Some cases of Addison's disease, it has been suggested, may be due to a pituitary defect, that is, to a deficiency of corticotrophin; improvement of patients with this disease has been reported to follow the administration of the adrenocorticotrophic principle.

Collip and his colleagues (1933) were the first to prepare from the anterior pituitary an extract which acted specifically upon the adrenal cortex, it contained minimal amounts of other pituitary principles. This extract was highly effective in restoring to normal the atrophied adrenal cortex of hypophysectomized animals. A protein with high corticotrophic activity was isolated by Li and his associates in 1942, and by Sayers and his colleagues in 1943. The material obtained by the latter observers is free from thyrotrophic and lactogenic actions. Its isoelectric point is at a pH between 4.7

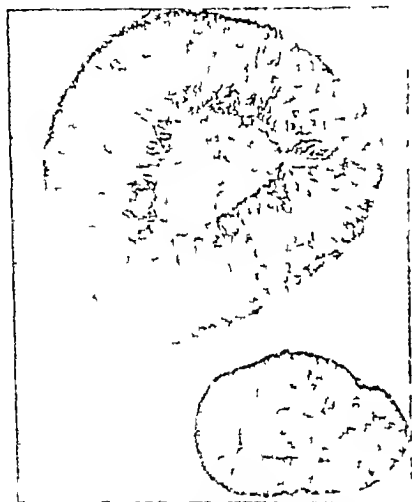


FIG 577 Showing adrenal atrophy following hypophysectomy. Normal gland above, gland of hypophysectomized animal on right. (After P. Smith.)

directly upon the thyroid cells. The effect of thyrotrophin in causing exophthalmos in the guinea-pig, and its probable role in the production of human exophthalmos is discussed in ch 58.

The normal rat or guinea pig, after an initial response period in which hyperplasia of the thyroid and a rise in the metabolic rate occur, becomes refractory to the thyrotrophic hormone. The refractoriness has been shown by Collip and Anderson to be due to the formation of some inhibiting substance (anti-hormone (see p 791)) which they have found in the serum of the treated animals. Friedgood also found that though guinea pigs responded at first to an extract of the anterior lobe of the pituitary by thyroid hyperplasia and a rise in the basal metabolic rate, the animals later be-

and 48 and its molecular weight is around 20,000. Li and associates later obtained a polypeptide from ACTH extracts by partial hydrolysis with pepsin or acid. The product of acid hydrolysis has a potency some 100 times greater than that of the original material, its molecular weight is around 1200.

Factors relating to the action and liberation of corticotrophin

Corticotrophin induces changes in the quantity and distribution of the doubly refracting (birefringent) lipid material in the cortical cells. Moderate doses cause a reduction of the fine dust-like particles. With larger doses the lipid is increased in the zona glomerulosa and zona reticularis and appears in the form of larger accumulations which, it is suggested by some observers, represent the cortical hormone or its precursor in the storage phase (p 794). Gratten and Jensen found that corticotrophin stimulated the production of those cortical principles possessing an oxygen atom at C₁₁ in the steroid molecule. The production of those hormones lacking an oxygen atom at C₁₁, it is thought, is not stimulated by this pituitary principle.

The basophil cells are probably the source of corticotrophin. A great reduction in the number of these cells is said to occur in some cases of Addison's disease.

The gonadotrophic hormones and prolactin are dealt with in chapter 61.

The regulation of ACTH secretion is described on pages 791 and 844.

The ketogenic effect

If an extract of the anterior lobe is injected into fasting rats, or rats receiving a diet of butter fat, a rise in acetone bodies in the urine and in the blood occurs (Burn and Ling, Anselmino and Hoffmann). The effect is not mediated through the thyroid for it is obtained in thyroidectomized animals. It was thought, however, to be dependent upon the adrenal cortex since adrenalectomized animals treated with a potent ketogenic extract do not show ketonuria (Fry), whereas removal of the adrenal medulla alone does not prevent the response. It has been found, however, that the effect of the adrenal cortex is only upon the excretion of ketone bodies, adrenalectomy does not prevent their rising in the blood. Ketogenic extracts also cause an increased deposition

of fat in the liver and a diminution of body fat, indicating the transference of fat from the latter to the former site, which suggests in turn that the ketosis is the result of the stimulation of fat catabolism, and the production of ketones beyond the capacity of the body to oxidize them. The ketogenic effect is closely associated with the growth hormone and the diabetogenic factor (see below). Animals develop a resistance to the action of ketogenic extracts after they have received a number of injections.

The insulin-antagonizing (glycotropic) and diabetogenic factors

(see also p 680)

Many observations associate the adeno-hypophysis of the pituitary with carbohydrate metabolism. (a) Hyperglycemia, glycosuria, lowered sugar tolerance, are among the effects of over-activity of the anterior lobe in man (see acromegaly, p 799). All the symptoms of a true diabetes mellitus may be present in this condition. (b) Evans and others observed that dachshund pups given daily injections of anterior lobe extract as well as increasing 100 per cent in weight, exhibited great thirst, polyuria and glycosuria, and had a fasting blood sugar of 232 mg per 100 cc. (normal 100 mg per cent). (c) Hypophysectomy renders an animal hypersensitive to insulin and resistant to the hyperglycemic action of adrenaline, while implantations of anterior lobe or injections of an anterior lobe extract reduce the hypoglycemic action of the pancreatic hormone. It was also shown by Houssay that the diabetic symptoms following pancreatectomy almost disappear after hypophysectomy, but return when anterior lobe tissue is transplanted into the depancreatized and hypophysectomized animal. A pancreatectomized and hypophysectomized dog (Houssay dog) will live for several months without the aid of insulin, during fasting the blood sugar is within normal limits, or there may be hypoglycemia, ketonuria is absent and the glycogen content of the liver is near the normal value. Houssay and Biasotti concluded that the anterior lobe was concerned in the production of sugar from endogenous protein (see also p 628). Corkill, Marks and White suggested that the glycogen reserves of the liver are less readily mobilized after hypophysectomy, and that this is a contributory factor in the hypersensitivity to insulin of animals subjected to this opera-

tion The hyperglycemic effect of anterior lobe extracts is not brought about through the adrenal since it occurs after adrenalectomy (Houssay)

In 1937 Young reported the extremely important finding that dogs given a series of daily injections of a crude extract of the anterior lobe became *permanently* diabetic (hyperglycemia, glycosuria and ketonuria), i.e. the diabetogenic effect continued after the administration of the extract was stopped, and persisted indefinitely. The pancreases of such animals show complete degeneration of the islets of Langerhans (Richardson and Young) and give an extremely low insulin assay (Best and Campbell). These changes probably represent an "exhaustion atrophy" which accounts for the permanent diabetic state induced by the extract. The diabetogenic factor should not be confused with the insulin antagonizing principle or, as Young calls it, the glycotropic factor. The latter exerts a true anti-insulin effect. Extracts have been secured, for example, which, though reducing or preventing the hypoglycemic action of injected insulin, do not raise the blood sugar when administered alone. The glycotropic factor which is distinct from the thyrotropic and gonadotropic principles and probably from the lactogenic factor, also inhibits the actions of insulin in accelerating the fall in blood sugar in hepatectomized animals (Himsworth and Scott) and in causing the deposition of muscle glycogen (Marks).

The parathyrotrophic effect

There is a considerable body of evidence which suggests that the anterior lobe exerts a controlling influence upon parathyroid function. The association in the human subject of parathyroid adenomas with pituitary tumors has been reported. Decalcification of the skeleton occurs in pituitary basophilism (p. 799), though in a case of this disease Cushing found degeneration rather than hyperplasia of parathyroid tissue. Atrophy of the parathyroids has been reported in dogs (Houssay and associates) and in rats (Smith) following hypophysectomy. Collip and his colleagues observed hypocalcemia and degenerative changes in the parathyroids in dogs after the removal of the pancreas and the hypophysis. Removal of the pituitary alone, however, does not lower the serum calcium, though a rise in serum calcium has been observed to result from injections of anterior lobe extracts. Anselmino and Hoffmann claim to have produced parathyroid hyperplasia, often accom-

panied by hypercalcemia, by the injection of an anterior lobe extract into animals, and postulate the existence of a specific parathyrotrophic hormone. Similar results have been reported by Hertz and Kranes. Finally, Hertz and Albright have reported that the urine of a subject in whom there was a general hyperplasia of parathyroid tissue (hyperparathyroidism, p. 858), when injected into rabbits caused hyperplasia of their parathyroids. Although these several observations suggest some relationship between the pituitary and parathyroid function, there is no conclusive evidence that a specific parathyrotrophic hormone exists, and that such a hormone is essential for parathyroid function is disproved by the fact that calcium metabolism, even in young animals, appears to suffer no impairment after hypophysectomy.

The pancreatrophic effect

Anselmino and Hoffmann reported the preparation of an anterior lobe extract which caused enlargement and an increase in the number of the islands of Langerhans together with a fall in blood sugar. The extract was free from thyrotrophic effects. Their results, they believed, indicated the presence of a pancreatrophic hormone which stimulated insulin production. Richardson and Young have furnished confirmatory evidence of a pancreatrophic effect. They observed a 100 per cent increase in the quantity of islet tissue in the pancreas of rats receiving daily injections of a crude anterior lobe extract; many mitotic figures were seen in the islet cells of dogs treated with the extract. Marks and Young found that in rats the administration of a pituitary extract increased the insulin content of the pancreas. Ham and Haist have obtained histological evidence of a stimulating effect of a pituitary extract upon the islet and acinar tissue. After a few daily injections mitotic figures appeared in the beta cells of the islets as well as in the acinar cells. The daily administration of extract over a longer period led to degranulation followed by hydropic degeneration of the beta cells and, eventually, atrophy of the insulin-secreting tissue.

The pancreatic effects of pituitary extracts which have been just cited do not permit one to conclude that a specific pancreatrophic hormone exists. Such effects more probably are brought about indirectly through the action of the diabetogenic or of the glycotropic factor, that is, by

increasing the demand for insulin. The fact that hypophysectomy is not followed by atrophy of the islets of Langerhans speaks very strongly against the effects being truly "trophic" in character.

THE CONTROL OF THE SECRETION OF THE ADENOHYPOPHYSIS

Since there is a lack of any clear evidence that the pars distalis receives fibers which might bring about a secretory response from its cells, it is very doubtful that this part of the hypophysis is under *direct* nervous control. Indeed in some classes of mammals, e.g., the whale, porpoise and armadillo, etc., the pars distalis is separated from the neural part by a fibrous septum which nerve fibers of the *hypothalamico-hypophyseal* tract cannot penetrate. Furthermore, section of the neural stalk, the only likely course which secretory fibers might take, often causes no noteworthy effects upon anterior lobe function and none which interference with the blood supply cannot explain.

Yet there are a great many observations from which it must be concluded that the activity of the anterior lobe is subject to nervous influences. The liberation of an "ovulation hormone" by the pituitary of the rabbit on sexual excitement (p. 886), the release of the adrenocorticotrophic hormone under states of emotional stress, as shown by Long and his associates, the secretion of thyrotrophic hormone upon exposure to cold (Uotila p. 738), and the effect of light (retinal stimulation) upon the sexual functions of birds and of many mammals are outstanding examples of a relationship between the output of anterior lobe hormones and nervous activity. It has also been found (Markee and associates) that, whereas direct electrical stimulation of the pars distalis of the rabbit in 15 hourly periods over a total time of 5 days failed to cause ovulation, stimulation of the tuber cinereum for only 3 minutes was successful. How are these observations to be reconciled with the statements in the preceding paragraph? The vascular supply to the pars distalis (p. 784) offers the basis for an explanation. We have seen that blood which has first come into association with nerve cells in the tuber cinereum, and probably with other hypothalamic nuclei as well, empties into the sinuses of the pars distalis. It is reasonable to presume that an almost unique vascular arrangement of this nature possesses physiological significance. The cells of the hypothalamic nuclei have every ap-

pearance of being secretory in function (p. 1025). There is now little doubt that these cells, as a result of impulses received from other parts of the nervous system, discharge a secretion into the capillary blood which is conveyed to the sinuses of the pars distalis through the *hypothalamico-hypophyseal* portal system of vessels. Thus, a hypothalamic hormone is brought into relation with the cells of the adenohypophysis which produce the trophic hormones. In other words, the neurosecretory cells and the portal system of vessels serve as functioning links between the nervous system and the endocrine cells of the adenohypophysis. The conception of neurosecretory cells is, of course, not new. The existence of such cells in the nervous systems of both invertebrates (insects, mollusks, worms, etc.) and vertebrates including the human hypothalamic nuclei has been well established. The secretion of such cells has even been demonstrated as droplets of an acidophil material in the axons of the nerve cells. In fish also, similar material has been traced all the way along the axons from the hypothalamic cells to the hypophysis (Scharrer and Scharrer).

Further evidence for the control of adenohypophyseal activity by means of neurohormones elaborated in the hypothalamus and carried to the cells which secrete the trophic hormones is (a) The adenohypophysis is unresponsive to direct electrical stimulation, whereas electrical stimulation of the hypothalamus (tuber cinereum) is effective, (b) The effect of section of the neural stalk in abolishing control of ACTH secretion is inconstant, but Harris has shown that failure of this operation to prevent ACTH liberation in response to various states of stress is due to rapid postoperative restoration of the hypophyseal portal circulation. If revascularization is prevented the hypophysis remains unresponsive to hypothalamic influence, (c) Grafts of hypophyseal tissue in the neighborhood of the hypophyseal portal circulation respond to hypothalamic stimulation; those placed in more remote situations and which, therefore, are not vascularized from the hypophyseal-portal vessels, do not respond (Harris).

ANTIHORMONES AND PROHORMONES

Collip and his colleagues demonstrated that certain hormones when injected cause the formation of a substance in the serum of the treated animal which exerts a specific antagonism to, and may completely abolish the effect of, the injected hormone. They thought that the inhibitory principle was not an antibody in the ordinary sense, i.e., an immune body developed to some constituent of the hormone acting as antigen, but was rather in the nature of an *antihormone*. A substance

of this character was first obtained by Collip and Anderson in 1934 in the course of experiments with the thyrotrophic principle of the anterior pituitary. Rats after repeated injections of the thyrotrophic hormone became refractory, subsequently failing to respond to large doses. They were still responsive, however, to the thyroid hormone. It was found that the serum of these refractory animals was capable of preventing the effect of the thyrotrophic hormone upon non refractory animals. To give another example of this antagonism hypophysectomized rats treated with purified growth hormone fail to respond to it after a period of five or six weeks. The serum of such animals neutralized the effect of the growth hormone upon other untreated hypophysectomized rats. Besides anti thyrotrophic and anti growth principles, inhibitory substances have been shown to exist for the ketogenic, lactogenic and gonadotrophic principles of the anterior pituitary, respectively, as well as for the anterior pituitary like principle (HCG, ch. 61), but not for the adrenotrophic, or parathyrotrophic principles, or for insulin, or the sex hormones.

Collip suggested that the inhibitory principles or antihormones were present in the blood of normal animals, each serving to "oppose" or "buffer" the action of the corresponding hormone and that a given endocrine effect occurring naturally in the body was dependent upon the ratio existing between the hormone and antihormone concentrations.

Collip's observations have been confirmed repeatedly by other experimenters, so that the production of substances antagonistic to certain pituitary principles when administered, is a phenomenon established beyond all doubt. But there has been a difference of opinion as to the nature of the inhibitory substances. The problem is one of extraordinary complexity for experimental investigation. From several later observations, however, e.g., that rats do not develop antihormones to implantation of anterior lobe tissue of their own species, and antagonodotrophic substances are not formed in the blood of women injected with human chorionic gonadotrophin, it is most probable that the effect of these so-called antihormones represents a form of antigen antibody reaction due to contaminating foreign (antigenic) protein. In other words, the antihormone effect is not evident, unless the hormone itself is derived from another animal, especially one of another species.

Augmentation of the action of a hormone after repeated injections has also been observed. The augmenting substance which develops in the serum of an injected animal is called a *prohormone*. The augmenting action may precede the antagonizing or antihormone effect.

THE NEUROHYPOPHYSIS

An active extract of the pituitary was first obtained in 1894 by Oliver and Schafer, its effect in raising the blood pressure was described. Howell

three years later showed that the posterior lobe alone contained the pressor principle. The extract has been used in medicine for many years under such commercial names as *Pituitrin*, *Infundin*, *Infundibulin*, etc., or as the official solutions of the British and United States pharmacopoeias (B.P. and U.S.P.).² Much of our knowledge concerning the pharmacological and physiological actions of the principle of the posterior lobe has been gained from experiments with this relatively crude material and although it has been largely displaced by the more purified fractions *pitressin* and *pitocin* (p. 795) an account of the actions of the whole posterior lobe extract will first be given.

The actions of pituitrin or other standardized pituitary extract

These fall into 5 main groups

(1) **CIRCULATORY**. The blood pressure is raised, the systemic arterioles and capillaries both undergoing constriction. Marked pallor of the skin results. An initial depressor effect, or a pressor succeeded by a depressor effect, may precede the main rise in blood pressure. A second injection of the extract shortly after the first usually causes a depression of the blood pressure (*inversion effect*). The depressor effect may increase with repeated injection, it is cardiac in origin as indicated below. Pituitrin causes constriction of the coronary and pulmonary vessels but dilates the cerebral and renal vessels. The dilator effect upon the two last mentioned sets of vessels is a passive one being caused by the rise in systemic blood pressure. The heart is slowed by pituitrin if the vagus nerves are intact, the effect being a reflex result of the blood pressure rise, but increased cardiac rate occurs if the nerves have first been cut. Some dilatation of the heart and weakening of its beat occur in the dog, rabbit and, with large doses, in the cat. The coronary constriction and the weakening effect upon the cardiac muscle cause a reduced cardiac output and a fall in pressure in the pulmonary artery (dog and rabbit). These experimental results indicate that commercial pituitrin is of no value *clinically* as a means of

² The official solution of U.S.P. (XIII) *Posterior Pituitary Injection* is an aqueous solution of the principles of the posterior lobe of the pituitary of "healthy domesticated animals used as food by man". It has a specific activity (oxytocic, antidiuretic or pressor) of 10 posterior pituitary units per cc. A *Posterior Pituitary Unit* (B.P., U.S.P., or International) is the potency of 0.5 mg. of the Posterior Pituitary Reference Standard.

strengthening the action of a failing heart but may actually exert a deleterious effect. The fall in blood pressure which follows repeated injections of pituitrin (inversion effect) is due, according to Melville and Stehl, to the weakening of the heart and not of vascular origin. The portal venous pressure is reduced by pituitrin—as a result of constriction of the splanchnic vessels.

(2) **PLAIN MUSCLE** is stimulated by pituitrin. This action differs from the smooth muscle stimulating action of adrenaline in that it does not parallel the action of the sympathetic nerves, smooth muscle, receiving motor innervation from the parasympathetic, is excited as well. The muscular walls of the *uterus* (*oxytocic effect*), *intestine*, *gall-bladder*, *ureter* and *urinary bladder* (destrusor and trigone) are excited. Peristalsis in the human small intestine and to a less extent in the colon, is stimulated, but the tone is unaffected. Sometimes the smooth muscle of the bronchioles is stimulated, but this is due to contamination of the extract with histamine and is not specific. The stimulating effect of pituitrin upon the isolated virgin guinea-pig's uterus was demonstrated by Dale in 1909, it is used as a means of assaying the potency of pituitary extracts. Some highly purified preparations exert an oxytocic effect in a dilution of 1 part in 2,000,000,000. The effect of pituitrin upon the uterine muscle is antagonized by the hormone of the corpus luteum (see p 880). The oxytocic effect varies with the species, and, as a consequence of the interplay of other hormones, especially of the luteal hormone, with the phases of the sexual cycle (e.g., period of estrus, pregnant or non-pregnant state of the uterus). An interrelation between the actions of adrenaline and pituitrin is indicated by the fact that if the non-pregnant uterus of the cat is first treated with pituitrin, adrenaline causes contraction instead of the usual relaxation (p 831). The human uterus is most sensitive to the extract at the end of pregnancy, and pituitrin is used as an obstetric aid to induce uterine contraction after the expulsion of the placenta and so to prevent or check post-partum hemorrhage. The posterior pituitary principle also causes a temporary increase in the secretion of milk in lactating animals (so-called *galactagogue action*). This is not, however, a specific secretory effect but is due simply to the stimulation of smooth muscle in the walls of the mammary alveoli and ducts, and the expression of pre-formed milk (ch 61).

Respiratory effects, e.g., increased rate of breathing, alternating at times with periods of apnea, are produced by pituitrin but they are secondary to the effects on the circulation.

(3) **THE ANTIDIURETIC HORMONE (ADH)** The action of posterior pituitary extracts upon renal function was discovered by Magnus and Schafer in 1901, but they described only a *diuretic* effect. The diuresis, however, is of brief duration and may be preceded by anuria due to ureteral spasm, this early diuretic action is due to the rise in general blood pressure, and the passive dilatation, thereby, of the afferent glomerular vessels. The important and specific renal effect is, however, an increase in the reabsorption of water and, as a consequence, a reduction in urinary flow. In anesthetized animals the specific antidiuretic action is absent, diuresis due to the vascular factors is the outstanding effect. Pituitrin postpones for several hours the diuresis induced in normal animals by water drinking, and reduces the polyuria of diabetes insipidus. Associated with the antidiuretic effect is an increase in the percentage of sodium and chloride in the urine and, as a result of a reduction in the tubular reabsorption of these electrolytes, an increase in their total excretion (see also ch 35). This effect upon the tubular reabsorption of chloride, though until recently thought to be due to the antidiuretic hormone, appears now, from the work of Dicker and Heller, to be caused by traces of pitocin in the ADH preparations employed. Zinc salts in minute amounts prolong and enhance the antidiuretic effect of pituitary extracts.

The antidiuretic principle is an essential factor in the maintenance of the water balance of the body, being secreted when the need for water conservation arises. For example, the quantity of the hormone in the urine, which reflects presumably its concentration in the blood stream, is increased in dehydrated states but decreased in hydremia (Gilman and Goodman). In man, it is through a more active tubular reabsorption rather than by a reduction in the filtration rate through the glomerulus that the kidney plays its part in conserving water.

It is only in mammals and birds that pituitrin exerts any effect upon tubular reabsorption. And, as compared with mammals, the antidiuretic action of pituitrin in birds is feeble. Only these two classes secrete a hypertonic urine, and their nephrons alone possess a thin segment of Henle's loop. It is thought, therefore, that this part of the

renal tubules is the site of action of the antidiuretic hormone (see also ch 35). However, pituitrin is not without effect upon the water exchange of amphibia and reptiles. Frogs, for example, immersed in water increase their body weight if injected with pituitrin (*Brunn effect*) as a result of the uptake of water through the skin, they also lose water less readily in a dry environment if treated with pituitrin. This effect⁴ is due to a principle in pituitary extracts other than ADH, it is closely associated with the oxytocic principle but does not seem to be identical with the latter. It is called the *water balance principle* and can be prepared from the pituitaries of mammals as well as from those of amphibians.

The secretion of the antidiuretic hormone appears to be controlled mainly by the supra-optic nucleus of the hypothalamus and, to a less extent, by the tuber cinereum. The experiments of Verney and his associates upon water diuresis (chap 35) indicate that the adequate stimulus to the nerve cells leading to the release of ADH is the osmotic pressure of the blood (actually of the extracellular fluid). The location in the central nervous system of the cells which respond to such an osmotic change (osmoreceptors) is not known, but it is reasonable to presume that they are situated in the hypothalamus and probably in the supraoptic nucleus.

That the liberation of ADH is brought about through nervous influences of a more remote origin is probable. Stimulation of the central end of the vagus, for example, has been reported to reduce the urine flow (release of ADH) which is accompanied by the appearance of an antidiuretic substance in the urine, but does not occur after hypophysectomy. The repeated stimulation of the retina by flashes of light, in rats, is said to cause diuresis (suppression of ADH). The inhibition of water diuresis by exercise or emotion (p 457) is another example of a nervous influence upon the liberation of ADH.

The experiments of Pickford suggest that the immediate stimulus to the endocrine cells may be acetylcholine released at the nerve endings.

(4) **METABOLIC EFFECTS** Reduced tolerance for sugar, diminution in hepatic glycogen, hyperglycemia, glycosuria and a fall in the basal meta-

bolic rate follow the injection of pituitrin. The effect of insulin is antagonized. That is, to say, the pancreatic hormone produces less effect upon the blood sugar and a greater amount is required to produce hypoglycemic convulsions if its administration has been preceded by an injection of pituitrin. Pituitrin also causes an increase in liver fat. The effects on carbohydrate and fat metabolism are probably non-specific, but due rather to some extraneous factor associated with the posterior lobe secretion, and therefore does not represent a physiological action.

(5) **MELANOPHORE-EXPANDING (-DISPERSING) PROPERTY** In the skins of many cold-blooded animals are peculiar cells with branching processes and containing mobile pigment granules whose movement toward the periphery of the cell or toward the center is under hormonal influence. Such cells have been given the general name of *chromophores*. Those containing black pigment (melanin) are known as *melanophores* and those with red or yellow pigment are called *erythrophores* or *xanthophores*, respectively.

Hogben and Winton showed that pituitrin caused the pigment granules in the melanophores of the frog to become dispersed throughout the bodies and branching processes of these cells. This results in darkening of the skin. Thus, the injection of a drop or so of a solution of pituitrin into a frog causes its skin to become almost coal black as a result of the melanophore reaction. On the other hand, after hypophysectomy, owing to the disappearance of the pituitary hormone from the circulation, the pigment granules gather near the center of the cells and the skin becomes pale. Hypophysectomized tadpoles have, instead of the usual dark-brown or green color, a silvery appearance. The changes in color which certain amphibia, reptiles and fish undergo in order to blend into the color of their surroundings is largely due to variations in the concentrations of the pituitary hormone in the blood or to the balance between the concentrations of the melanophore-expanding principle and adrenaline. Blinding a frog deprives it of this adaptive power.⁵ Nervous

⁴ It is not seen in fish or reptiles, but in the latter, nevertheless, though not in the former, pituitrin exerts a renal effect, but one which is different from that which it induces in mammals, namely, a reduction of the filtration rate.

⁵ The phenomenon of light influencing structure and function through the mediation of a retinohypophyseal mechanism is not peculiar to cold blooded animals. The work of a number of experimenters indicates that the well known association of the seasonal periods with morphological changes (e.g., color and texture of hair or plumage) and with the sexual cycles of certain mammals and birds as well as with the migration of birds, is due, in part at least, to light acting upon the

impulses arising in the retina are believed, therefore, to govern the liberation of the melanophore expanding principle. The *production* of melanin in the skin of the frog is said to be stimulated by this principle.

In most animals, the melanophore-expanding principle is elaborated by the pars intermedia.⁶ Cultures of tissue from the pars intermedia alone yield this principle, whereas, cultures of the neural part of the pituitary or of the anterior lobe do not. Nevertheless, the principle finds its way into the neural part of the pituitary and unpurified extracts of the latter always contain it as a contamination. But it is possible to prepare it free from anti-diuretic, pressor and oxytocic effects. Moreover, the melanophore dispersing action of extracts from different parts of the posterior lobe do not run parallel with the pressor, oxytocic and antidiuretic properties. These facts argue for the melanophore effect being due to a separate and distinct hormone. Zondek has given the name *intermedin* or the *chromatophorotropic hormone* to this principle. It has been suggested that the melanophore-expanding principle also causes the migration of the melanin granules in the pigment layer of the retina (ch 73) and is, therefore, a factor in dark adaption of the eyes of higher vertebrates, but the experimental evidence in respect to such a function is conflicting. Nor is there any definite evidence that any relationship exists between this hormone and the occurrence of retinitis pigmen-

anterior pituitary through the medium of retinal impulses. Rowan, for example, was able to induce sexual activity in crows and canaries at any desired time of the year by varying the periods of their exposure to artificial light, it is also well known that the domestic fowl will lay regularly in winter if its period of exposure to light is lengthened by artificial means. Cognate experiments have been carried out upon mammals—ferrets, hedgehogs and racoons. The seasonal shedding of hair in the ferret has been shown to be conditioned by the length of the day, and optic nerve section or hypophysectomy abolishes the phenomenon. The sexual cycles are abolished, of course, by hypophysectomy (lack of gonadotropic hormone) but their periodicity in relation to the length of exposure to light is lost after optic nerve section alone. The extreme sensitivity to light of the pigeon in respect to the time of egg laying is extraordinary. The pigeon lays a pair of eggs in the morning, the second of the pair being laid with the regularity of clock-work half an hour after the first. Lengthening or shortening the period of exposure to light alters the duration of the interval between the laying of the first and second eggs of the pair.

⁶ In those species such as the chicken, porpoise and whale which do not possess a pars intermedia, the melanophore-dispersing principle is found in extracts of the anterior lobe, none is present in the posterior lobe.

tosa (p 805). It is natural that it should also be suspected of playing a rôle in cutaneous pigmentation in the human subject but again no reliable observation has been reported which might connect this principle with either normal or abnormal pigmentation in man.

Fractionation of the posterior lobe extract Pitocin and pitressin

Considerable quantities of histamine may be present in commercial pituitary extracts which have not been carefully purified, indeed it was believed for a time by some (Abel and associates) that the plain muscle stimulating and depressor effects were due simply to histamine. Abel also maintained that the other effects (pressor and antidiuretic) were due to a single hormone. The work of Dudley, and especially of Kamm and his associates has shown that the specific effects are not due to histamine, and that there are at least two distinct active principles in a posterior lobe extract. From the crude extract Kamm and his associates have isolated two relatively pure fractions which are called, respectively, *pitocin* and *pitressin*. These are white amorphous powders freely soluble in water. Chemical analyses of highly purified preparations have revealed a high amino-acid content of both pitocin and pitressin; cystine, tyrosine and arginine are found in great amounts. Their molecular weights are between 500 and 2000. Pitocin is the uterine-stimulating (oxytocic) principle in the official pituitary solution (p 792) but is between 100 to 350 times more potent than the latter. Pitocin acts upon the muscle of the uterine body but not upon the cervix.

Pitressin is responsible for the cardiovascular, intestinal, and antidiuretic effects of the official pituitary preparation. If a series of injections of pitressin be given at short intervals a certain degree of tolerance becomes established, the hypertensive effect being less with successive doses. It is possible that the antidiuretic action of preparations of pitressin is due to a separate principle, for the pressor effect is more readily destroyed by heat than is the antidiuretic action. Preparations of pitressin with a pressor activity 300 times greater than that of a solution of Standard Pituitary Powder have been prepared.

Pitocin exerts its greatest effect upon the human uterus in the later months of pregnancy. In the first few months this principle exerts no oxytocic action, but a slight effect upon the uterus, probably of vascular origin follows the injection. ◻

pitressin at this time. In the foal, pitocin causes a fall in blood pressure. These fractions of posterior lobe extracts are not absolutely pure in their effects, pitocin has a slight pressor action and pitressin slightly oxytocic. But from a practical point of view a great advantage has been gained by the separation of these two principles in relatively pure form, for the effect upon the uterus can now be secured with little or no rise in blood pressure which it is often advisable to avoid.

Both pitressin and pitocin cause hyperglycemia, but which one will have the greater effect depends largely on the species. In rabbits pitressin has the greater hyperglycemic action, whereas in dogs, pitocin is more effective. The rise in blood sugar induced by these principles is due to the breakdown of glycogen (glycogenolysis) in the liver.

The nature of pitocin and pitressin from a purely physiological point of view, i.e., whether or not they represent distinct hormones secreted by the neurohypophysis, is in doubt (see p. 784).

Van Dyke and associates have isolated a non-crystalline material of protein nature from posterior lobe extracts which, from electrophoretic, ultracentrifuge and solubility studies, seems to be a pure substance free from contaminating material. It possesses oxytocic and pressor activities in the same ratio as in the extracts.

Functions of the neural lobe

Though a posterior lobe extract when injected produces very definite effects, such might not represent the physiological actions of the gland within the body. In other words, the effects following injection might be due to substances which, though obtainable by extraction, were merely incidental to the chemical procedures, i.e., artifacts, and not principles of a true internal secretion. Proof that any given effect was hormonal would be the demonstration that removal of the gland produced a contrary effect which could be corrected in turn by injection of the extract. There is little definite evidence, for example, that in mammals the posterior pituitary secretes a pressor substance which maintains the normal blood pressure, and little to suggest that it is concerned with capillary tone, as was suggested by Krogh from experiments upon the frog. It is true that in the latter species and in the toad the posterior lobe does appear to be essential for the maintenance of capillary tone. Dilatation of the cutaneous vessels occurs in these animals after total hypo-

physectomy but not after removal of the anterior lobe alone.

The isolation of an oxytocic principle from the posterior lobe led, as it was bound to do, to the theory that the elaboration and discharge of such a principle by the pituitary constituted an important factor in the birth mechanism, serving to stimulate the uterine contractions and expel the fetus at the end of the period of gestation. This conception of posterior lobe function, attractive and plausible though it was, appeared to be refuted by certain observations. Allan and Wiles, for example, reported that normal parturition occurs in hypophysectomized cats, and Selye, Collip and Thomson found that though the period of gestation in rats was prolonged after hypophysectomy, no abnormality in the birth mechanism was detectable. Bell and Robson demonstrated an oxytocic principle in the blood of pregnant cows, women in labor and in non-pregnant rabbits, but since there was no relation between the concentration of this substance and any period of gestation, concluded that it was not concerned in the initiation of labor.

Such negative findings aroused scepticism as to a physiological relationship between the pituitary and the parturient uterus. However, this uncertainty has been largely dispelled by recent work. Haterius and Ferguson, experimenting with rabbits from 2 to 80 hours *post partum*, observed increased frequency and amplitude of the uterine contractions upon electrical stimulation of the pituitary stalk. The response closely resembled that induced by pitocin and was obtained after spinal transection, vagotomy and severance of both splanchnic nerves. It was abolished by destructive lesions placed in the infundibulum. These results strongly suggest that the oxytocic principle of the pituitary is a true hormone. The continuous activity of the uterus in the puerperal rabbit appears in the light of the results of later experiments of Ferguson to be due to pituitary secretion initiated reflexly from the pelvic viscera, for the contractions undergo progressive diminution in amplitude after interruption of the afferent pathways, by section of the pituitary stalk or of the spinal cord above the 1st lumbar segment. Fisher and his colleagues had also found previously that after section of the hypothalamico-hypophyseal tract in cats the expulsive movements of the uterus at term were extremely sluggish. We know that this operation reduces almost to the vanishing point the active principles of the posterior pituitary.

Other significant facts lending support to the conception of posterior pituitary function just outlined are, the insensitivity of the uterine cervix to pitocin, and the influence of the female sex hormones upon the action of the oxytocic principle, namely, the depression of uterine sensitivity to the latter induced by progesterin and the greater sensitivity conferred by estrogen (p 880) Very suggestive in this connection, also, is the discovery of Cohen and Marrian (p 878) that the estrogenic hormone is present in the blood and excreted in the urine as an inactive compound (glycuronide of estriol) in pregnancy until just before the onset of labor when it appears in the free (active) form

Expansion of the melanophores in amphibia or of the erythrophores in the minnow is most certainly a hormonal effect, as is also the *antidiuretic action*

But whether the different hormonal effects obtainable with posterior lobe extracts, or with the fractions pitocin and pitressin, represent naturally occurring distinct chemical principles or different actions of a single large hormone molecule is uncertain Pitocin and pitressin may be merely chemical groups split off from such a parent molecule by the laboratory procedures—i.e., the products of a "chemical dissection" Evidence, none of which is conclusive, can be cited for either view These statements apply only to the principles of the pars nervosa, and not to intermedin which, it is generally accepted, is a separate and distinct hormone formed by the pars intermedia

THE CONTROL OF THE ACTIVITY OF THE NEUROHYPOPHYSIS AND THE SOURCE OF ITS HORMONES

While the evidence is against the adenohypophysis being under direct nervous control, the clear demonstration of nervous pathways to the neural lobe, the depletion of posterior lobe hormones after section of the hypothalamico-hypophyseal tract, as well as the experiments already mentioned in respect to the liberation of the oxytocic and melanophore expanding principles, leave little doubt that this part of the pituitary is controlled directly by nerves

What cells of the neurohypophysis furnish the secretion is a question which has engaged the attention of many investigators The earlier work of Herring, of Cushing and Goetsch and others suggested that the cells of the pars intermedia and pars tuberalis elaborated the hormone or hormones, and that the hyaline bodies (p 783) seen in these regions represented ripened cells filled

with secretion Geiling and Oldham, on the contrary, are convinced that the pressor, oxytocic and antidiuretic hormones are formed, not in the pars intermedia, but in the neural lobe itself One would think that this must be so, for the pars intermedia is absent from the pituitary of the whale, porpoise and chicken, animals in which there is no lack of the antidiuretic hormone Nor can ADH be extracted from the neural lobe of ordinary laboratory animals if part of the pituitary is atrophied though the pars intermedia appears normal Since the discovery of neurosecretory cells the absence of tissue of a glandular nature from this part of the pituitary is no obstacle to the acceptance of such a view and, as Geiling points out, the adrenal medulla is of neural origin Another reason for believing that the neural tissue, probably the pituicytes, elaborates the posterior pituitary principles is afforded by the experiments of Fisher and his colleagues already referred to After section of the nerve fibers of the neural stalk degenerative changes confined to the posterior lobe appear, and the pressor, oxytocic and antidiuretic principles are greatly diminished, whereas the pars intermedia contains the normal amount of melanophore-expanding principle Most students in this field believe that the hyaline bodies described by Herring are artifacts or at any rate do not represent the posterior lobe secretion but whatever their nature and function may be, an experiment performed by Maddock suggests that they traverse the posterior lobe and infundibular stalk to be discharged into the third ventricle When a compressing clip was placed upon the neural stalk numbers of hyaline bodies gathered, as though blocked, on the pituitary side of the clip Injected particles of India ink have also been seen to traverse the neurohypophysis and stream up the stalk, which strongly suggests that this is the pathway taken by the posterior lobe hormones Hyaline bodies in the act of discharging into the ventricle have been described Detection of posterior pituitary principles in the cerebro-spinal fluid by means of the oxytocic and melanophore tests have been reported by some workers, others, however, have failed to confirm the observation The recent experiments of Zondek indicate that intermedin leaves the pituitary via the infundibular stalk This principle was recovered from the walls of the third ventricle in the region of the autonomic centers but from nowhere else outside the pituitary

Circumstantial evidence for the view that the

posterior lobe hormone is discharged into the ventricular system has been furnished by Cushing. He found that the injection of commercial pituitrin into the lateral ventricle of man produced the most striking effects, which commenced within five minutes of the injection. These effects resemble in several respects those characteristic of parasympathetic stimulation, and are quite different from those resulting from the intramuscular injection of pituitrin. They include (a) flushing of the skin and most profuse sweating, (b) fall in rectal temperature from about 99.0° to 94.0° , (c) nausea, vomiting and contractions of the small intestine, (d) rise in blood pressure after an initial fall and some increase in heart rate, (e) salivation and lacrymation, (f) fall in metabolic rate (from 13 to 31 per cent) within $1\frac{1}{2}$ hours. These effects are not due to histamine. Comparable effects have been produced by Fulton in the monkey. The intraventricular injection of pilocarpine (a drug which stimulates structures innervated by parasympathetic nerves) causes comparable effects. Atropine, which abolishes parasympathetic effects, prevents the customary results following the intraventricular injection of either pituitrin or pilocarpine. Furthermore these agents have not their usual action in persons in whom the tissue of the hypothalamic region has been destroyed by disease, or in persons under the influence of an anesthetic such as avertin, which is believed to act upon the hypothalamus. Penfield has reported manifestations similar to those described above in a patient with a lesion (tumor) in the region of the hypothalamic nuclei; he refers to the condition as *diencephalic autonomic epilepsy*. On the basis of his results Cushing conceives that the secretion of the posterior lobe passes into the cerebro-spinal fluid of the ventricle and diffuses through the nervous tissue to act upon the parasympathetic center in the hypothalamus (tuber cinereum). On the other hand, the secretion of the posterior lobe is controlled by impulses arising in higher nervous centers and relayed to the gland through the hypothalamic nuclei (see p. 1025). It is altogether likely that the osmotic pressure of the blood, which we have seen is a factor in the control of ADH liberation, does not act directly upon the neural lobe, but upon the cells (osmoreceptors) of the hypothalamus (see also p. 605 and chapter 35). This conception of a nervous-hormonal synergism finds its analogy in the sympatho-adrenal system in which impulses flowing along

sympathetic pathways cause the liberation of adrenal secretion, which in turn reinforces the actions of the sympathetic.

DISORDERS OF THE PITUITARY IN MAN

Derangements of pituitary function may take the form of overactivity or of deficiency. In the former case tumors composed of functioning endocrine tissue are frequently the cause of the disorder, in the latter, atrophy or degeneration of the specific secreting cells, either primarily or as the result of mechanical pressure by tumors, may be responsible. A pituitary tumor of the anterior lobe may be composed of any of the cellular elements of the gland—*chromophobe*, *acidophil* or *basophil* adenomas. Squamous-celled growths (craniopharyngeomas) may also arise from epithelial rests—remnants of Rathke's pouch—near the root of the infundibular stalk. As a result of the confined position of the pituitary within the sella turcica the entire gland is likely to suffer from pressure effects when one of its parts becomes enlarged. For this reason and on account of the proximity of other important structures, e.g., hypothalamus and optic chiasma, and the proclivity of tumors to invade or press upon neighboring structures, the manifestations of a pituitary tumor are not always referable simply to the part of the pituitary originally involved. Any function—growth, sex, water elimination, or the metabolism of carbohydrate or fat, presided over by the pituitary-hypothalamic mechanism may therefore be disturbed by a lesion in this region. Nevertheless, the site wherein the tumor arises and the nature of the cells of which it is composed do very often determine the predominant features of the condition, and certain fairly well defined groups of symptoms (syndromes) are recognized. To these may or may not be added symptoms referable to pressure upon, or to irritation or destruction of, near-by nervous structures.

The following is a short classification of pituitary disorders—

Anterior lobe	{ overactivity	Acromegaly
		Giantism
	{ deficiency	Pituitary basophilism—
		Cushing's disease
	{	Dwarfism
		Pituitary cachexia—
		Simmond's disease
		Acromicria

Posterior lobe deficiency or hypothalamic lesion	Diabetes insipidus
Anterior lobe deficiency together with posterior lobe deficiency or hypo- thalamic lesion	Dystrophia-adiposo geni- talis (Fröhhch) (a) Infantile or juvenile type (b) Adolescent or adult type

ACROMEGALY

This condition was first described by Pierre Marie in 1885. It is due to the excessive elaboration of the growth hormone (p 785) during adult life, i.e., after the usual age of full skeletal growth (fig 578). An adenomatous tumor of the anterior lobe composed of acidophil cells is responsible for the hypersecretion. The characteristic features of the condition are (a) overgrowth of the bones of the hands, feet and face. Of the latter, the mandible, nasal bones and supraorbital ridges are especially involved (fig 579). The feet and hands are greatly increased in size, the latter being usually broadened and the fingers thickened, under the X-rays, the terminal phalanges appear tufted, thus resembling a wheat-sheaf in shape. Bowing of the spine (kyphosis) is common. The soft tissues of the nose, lips, forehead and scalp are thickened, the latter being thrown into folds or wrinkles (bulldog scalp). There is a general overgrowth of body hair. (b) Atrophy of the gonads and suppression of the sexual function (amenorrhea in women, impotence in men). (c) Moderate increase in the urinary excretion of 17-ketosteroids and corticoids with reduced excretion of gonadotrophins. In the earlier stages, however, there may be evidence of increased sexual function. (d) Enlargement of the viscera (splanchnomegaly). The tongue, lungs, thymus, heart, liver and spleen are greatly enlarged. The thyroid, parathyroids and adrenals may show hypertrophy of adenomatous growths. Hyperthyroidism may be present in the early stages. (e) Glycosuria and hyperglycemia are common, and a condition indistinguishable from diabetes of pancreatic origin may be present. The metabolic rate may be raised by from 10 to 70 per cent, the specific dynamic action of protein is not altered (p 640).

GIANTISM

Giantism is due to a pituitary lesion of a similar nature to that responsible for acromegaly, but the condition arises in pre-adult life, i.e., before ossification is complete (fig 5710). A general over-

growth of the skeleton results and the production of persons of enormous stature—7 or 8 feet or more in height. The limbs are usually disproportionately long. The viscera are not enlarged out of proportion to the frame unless, as is sometimes the case, the giantism is accompanied by the characteristic features of acromegaly, as may occur after adolescence.



FIG 578 Acromegaly, together with enlarged stature—acromegalic giantism (After Cushing and Davidoff)

CUSHING'S DISEASE CUSHING'S SYNDROME

This is a rare disease, its main features are (a) Obesity of the trunk (especially of the abdomen), face and buttocks, but not of the limbs, these latter show some wasting. The fattening of the face leads to rounding of the facial contours which obscures the bony structure, producing the so-called "moon face". There is thus a re-distribution of body fat, which is mobilized from the limbs and deposited in the regions mentioned. The fatty deposits are frequently tense, tender and painful. Purplish striae, due to distention, are present over the lower abdomen. (b) Polycythemia with cyanosis of the face, hands and feet, pigmentation of the skin and excessive growth of hair.

Women may grow a mustache or a beard (c) Loss of mineral from the bones, leading to osteoporosis, softening or brittleness The softening often involves the dorsal vertebrae, and causes kyphosis, which with the deposition of fat in the interscapu-

and hypertrophy of the adrenal cortex with signs of hypersecretion of its hormones (h) Increased urinary excretion of 11-oxycorticosteroids, and often of 17-ketosteroids, high uric acid/creatinine ratio (i) Eosinopenia, lymphopenia, low blood

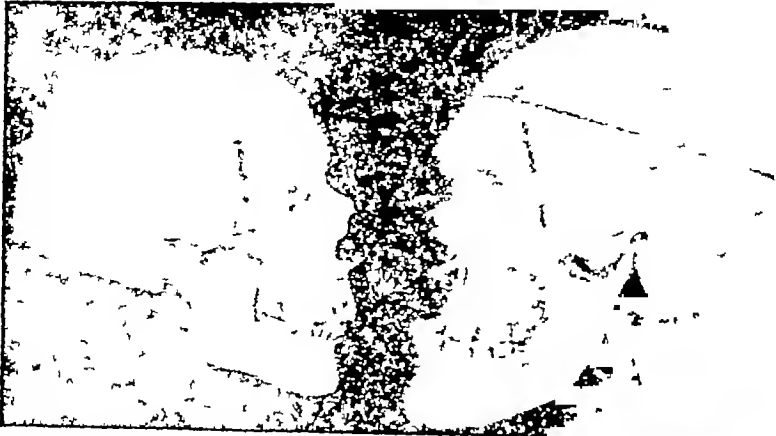


FIG 579 Skulls of a normal person (left) and of an acromegalic (right) After Cushing and Davidoff)



FIG 5710 Gigantism

lar region gives the appearance known as "buffalo neck" (d) Systolic hypertension (e) Suppression of sexual functions (f) Hyperglycemia and glycosuria and in some cases increased urinary excretion of nitrogen (g) Atrophy of testes or ovaries,



FIG 5711 On left, at 20 years of age, right, 5 years later at height of the disease. (From Cushing, after Turney)

potassium and retention of sodium chloride with increase of body fluids and a tendency to edema

The disease as originally described by Harvey Cushing (1932) was associated with a basophil adenoma of the adenohypophysis⁷ It has since

⁷ Susman found, however, that in the postmortem examination of a large number of pituitaries, small basophil adenomas were present in 3.1 per cent, though no signs of basophilism was observed during life

been found that the syndrome may arise either as a result of a pituitary tumor (excess ACTH production and stimulation of the adrenal cortex) or from primary hyperplasia or tumor of the adrenal cortex. The manifestations, it will have been observed, are in either case referable for the most part to excessive adreno-cortical secretion. Sometimes Cushing's *disease* is restricted to cases associated with a pituitary tumor, and Cushing's *syndrome* applied to the condition due to primary adrenal tumor or hyperplasia.

DWARFING

The arrested skeletal development which results from deficiency of the growth hormone of the anterior pituitary is spoken of as the *Lorain* type of infantilism (fig 57 12). These dwarfs are usually, though not invariably, undeveloped sexually. They do not show deformity or as a rule mental inferiority, and are generally not unattractive in appearance. Sometimes, however, they are wizened and except for their miniature stature appear older than their years (progeria), then the condition may be considered the counterpart of Simmonds' disease, but commencing before puberty. The anterior lobe dwarf at adult age may be no more than 3 or 4 feet in height. During infancy and childhood, the ossification centers as observed by radioscopy appear normal and dentition is not delayed. The relative proportions of the different parts of the skeleton are not far from normal though they tend toward those characteristic of childhood, the head being large relatively to the body. Some encouraging results have been reported from the treatment of this type of dwarfism during early childhood with anterior pituitary preparations.

PITUITARY CACHEXIA (SIMMONDS' DISEASE)

This rare disease was first described by Simmonds of Hamburg. It is due to atrophy or degeneration of the anterior lobe. The main features of this disease form a picture which may be best described as that of a premature and rapidly developing senile decay (fig 57 13). They are largely referable to severe depression of activity of the adrenal cortex and other target organs, e.g., gonads and thyroid, as a result of deficiency of the trophic hormones of the pituitary.

Crook also states that a basophil adenoma or a general increase in basophil cells is an inconstant finding in Cushing's syndrome, but that hyaline degeneration of these cells (with or without adenoma) is invariable.

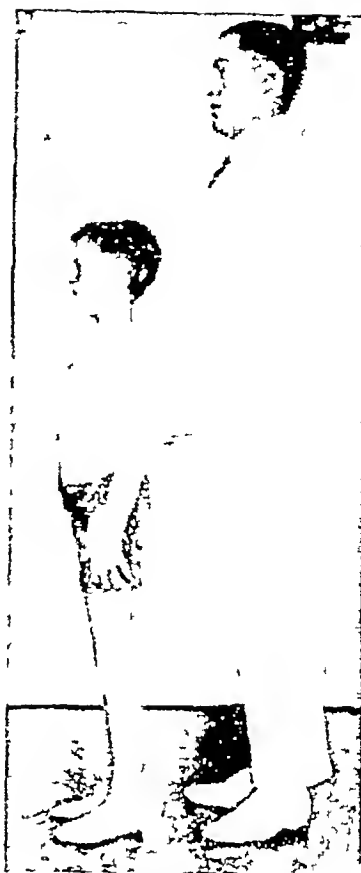


FIG 57 12 Infantilism due to anterior lobe deficiency. Age 21 years. Man on right, 5 feet 7 inches. (After Lissner.)



FIG 57 13 Pituitary cachexia. Photograph on left patient aged 42 years, photograph on right, the same woman at the age of 34 years. (From Zondek, *Diseases of the Endocrine Glands*, Arnold, London.)

(a) *General appearance*—loss of axillary and pubic hair, hair of the head prematurely gray and sparse, loss of teeth, skin of face wrinkled and dry, often great emaciation, smallness of hands and feet and shrunk appearance of the facial features. (b) *Anemia*. (c) *Low*

metabolic rate and hypoglycemia (d) Amenorrhea or impotence (e) Mental deterioration, muscular weakness, death in coma (f) Atrophy of the gonads and a general smallness of the internal organs (*splanchnomicroia*) (g) Reduced urinary excretion of gonadotrophins, 11-oxy-corticoids, and 17 ketosteroids, and depression of protein bound iodine of the blood (h) Acute adrenal insufficiency may occur

The disease most commonly follows childbirth in which there has been severe hemorrhage with peripheral circulatory collapse, necrotic areas are found postmortally which are attributed to the resulting ischemia of the gland. Mulder manifestations of the disease developing after childbirth have been described by Sheehan, which are now referred to as *Sheehan's syndrome*.

DIABETES INSIPIDUS

Diabetes insipidus is a condition in which large quantities of urine of very low specific gravity, 1.002 to 1.006, and low chloride content, are excreted. In an ordinary case the daily output of urine is 4 or 5 liters, but daily amounts several times these figures have been reported. A corresponding increase in the fluid intake occurs and thirst is often intense. The condition frequently accompanies tumors of the pituitary or hypothalamic region. It has been ascribed to posterior lobe deficiency, since it is relieved by injections of pituitrin and in earlier experiments upon animals was a common result of posterior lobe ablation. It has been shown, however, by Bailey and Bremer, and later by others, that in animals puncture of the hypothalamus in the region of the tuber cinereum without any apparent injury to the pituitary causes polyuria. Diabetes insipidus also ensues in conditions (epidemic encephalitis) involving the hypothalamic region. On the other hand, the aforementioned antidiuretic effect of pituitrin and the experiments of Maddock, in which the condition was induced in animals by the application of a clip to the pituitary stalk, point to deficiency of the posterior lobe of the pituitary as a factor. It is more correct, however, to look upon diabetes insipidus as the result of a disorder affecting the integrity of the hypophyseal-hypothalamic mechanism rather than as being always dependent upon either the pituitary or the hypothalamus exclusively. Involvement of one or other element, or of the nervous connections between the two would therefore cause the disease. Fisher, Ingram and Ranson found, for example, that a lesion inter-

rupting the supra-optico-hypophyseal tract of nerve fibers resulted in diabetes insipidus together with atrophy of the cells of the supra-optic nucleus and of the neural lobe of the pituitary. The atrophic posterior lobe was found to be almost devoid of antidiuretic, oxytocic and pressor principles.

It will be recalled (chap. 35) that only about 15 per cent of the glomerular filtrate is reabsorbed by the nephron distal to the proximal convoluted tubule ("facultative" reabsorption) and upon this fraction alone is the action of the antidiuretic hormone exerted, the reabsorption of water by the proximal tubule ("obligative" reabsorption) is not altered. Therefore, a limit is set to the increase in urine output even though ADH be completely absent. If, say, the glomerular filtrate amounts to 180 liters daily, at least 150 liters must be reabsorbed under all circumstances, leaving less than 30 liters as the maximum amount of urine which could be produced.

The failure of total hypophysectomy to produce diabetes insipidus is due, according to the widely accepted theory of von Hann, to the secretion by the anterior lobe of a hormone which antagonizes the post-pituitary antidiuretic factor, i.e., to the production of a *diuretic* principle. Von Hann based his view on the postmortem records of twenty subjects with lesions involving the posterior lobe. In nine, no functioning anterior lobe tissue remained and these had not suffered from diabetes insipidus. The experimental results of Richter accord with these findings. Total hypophysectomy in rats was in no instance followed by permanent diabetes insipidus, whereas this condition was invariably produced by removal of the posterior lobe alone. This theory implies that in the control of the excretion of water, the adenohypophysis and the neurohypophysis play antagonistic roles. The balance normally established between the function of the two parts may be disturbed by, (a) removal of the neurohypophysis, (b) by interruption of its nerve connections with the hypothalamus, or (c) by the injection of an anterior lobe extract after total hypophysectomy. It has been found that an intact thyroid gland is necessary for the experimental production of diabetes insipidus, the condition once established is abolished by thyroidectomy, but can be induced again by thyroid administration. An experiment performed by Keller suggests that the diuretic action of the anterior lobe is brought about through its thyro-

trophic hormone When anterior lobe extract was administered daily to a hypophysectomized animal the water intake did not show any increase until the lapse of from 24 to 48 hours, it then in-

animal to continued treatment suggest very strongly the action of the thyrotrophic principle In diabetes insipidus the ingestion of large quantities of water of a temperature below that

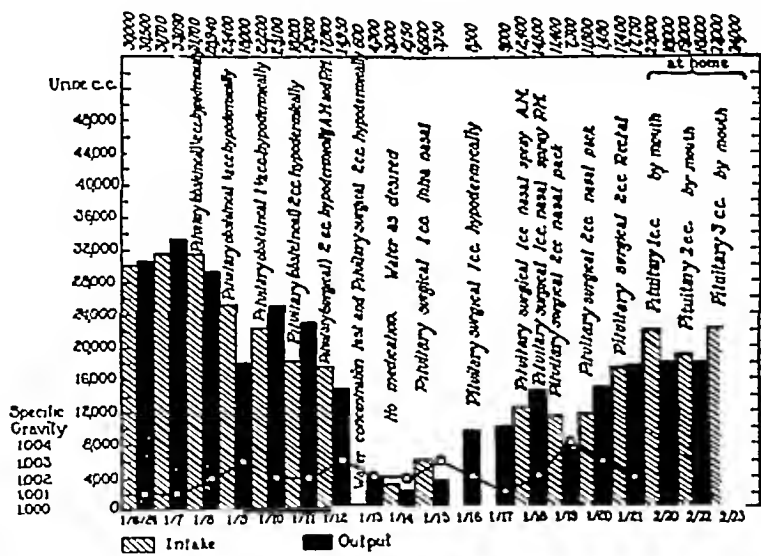


FIG 57 14. The effects of hypodermic injection and oral administration of pituitary extract in a severe case of diabetes insipidus (After Rowntree and Snell)



FIG 57 15 Male aged 8 Hypopituitarism following whooping cough, with characteristic growth and sex defects and obesity, Fröhlich's syndrome (After Gardiner-Hill)

creased progressively for 12 days and from then on gradually declined to the normal level, though the injections were continued The delay in onset and gradual development of the polydipsia together with the apparent refractoriness of the

of the body necessitates readjustments in the heat-regulating mechanisms Vasoconstriction of the skin vessels occurs, the extremities are often cold and may be cyanotic, and the subject is very sensitive to cold The vasoconstriction is apparently a compensatory measure for heat conservation Shivering and a rise in the metabolic rate may also result, the food intake is then, as a consequence, increased

Pituitrin administration is the only available means for the control of diabetes insipidus (fig 57 14) and, as mentioned below, this is not successful in some cases Pituitrin (or the pitressin fraction) may be given by injection or by means of pledgets of absorbent cotton soaked in the solution and inserted into the nose The polyuria of nephritis is not influenced by pituitrin administration It has been pointed out by Snell and Rowntree that diabetes insipidus following epidemic encephalitis (with hypothalamic involvement) is little influenced by pituitrin injections, and the thirst to a less extent even than the polyuria In these cases, therefore, pituitrin injections may result in such a disturbance of the water balance as to cause water intoxication, the latter may also occur in an ordinary case of diabetes insipidus if pituitrin is administered and the patient requested to maintain his usual fluid intake The urea clearances are normal as a rule in diabetes insipidus

The fundamental mechanisms concerned in the polyuria of diabetes insipidus appear to be allied in nature to those underlying the diuresis of water drinking in normal persons. But the thirst and the consequent excessive ingestion of water ex-



FIG 57 16 Extreme case of obesity, due to pituitary or hypothalamic disease. (After Timme.)



FIG 57 17 Laurence Biedl Moon syndrome. Note the presence of six toes (After Weiss)

perienced by subjects of diabetes insipidus is not primary, as in the water diuresis of normal persons, but is secondary to the polyuria and the resulting tendency toward dehydration the deprivation of water carried to the limit of endurance does not prevent the excretion of large quantities of water

The results of Richter's work upon rats, in which diabetes insipidus was produced by hypophysectomy or a stab wound in the floor of the third ventricle, indicate that in the experimental condition polyuria is also primary. He states that polyuria preceded polydipsia and that animals deprived of water continued to excrete large quantities of urine.

About 10 per cent of cases, classed as diabetes insipidus, do not respond to postpituitary preparations, or do so very poorly. Such cases are congenital or show a familial tendency, and are thought to be due to an inherent defect of the renal tubules with respect to the reabsorption of water—a persistence, it would seem, of the infantile incapacity of the tubules to concentrate the urine.

Polyuria, simulating in some ways diabetes insipidus, occurs in *chronic nephritis* (ch 36) and in *psychogenic polydipsia*. The polyuria in chronic nephritis is not corrected by pituitary extracts and in psychogenic polydipsia thirst is *primary*, the urine volume is reduced and the specific gravity of the urine is raised by rigid deprivation of water, the polyuria is simply a normal diuretic response to water drinking.

DYSTROPHIA ADIPOSEO-GENITALIS

Dystrophia adiposo-genitalis, as its name suggests, is a condition in which obesity, sexual infantilism and dwarfing (if the condition occurs during the growing period) are the essential features. It is due to a lesion of the anterior lobe of the pituitary, which accounts for the sexual immaturity and dwarfing, and of the hypothalamus (or posterior pituitary). The experiments of Smith with hypophysectomized rats and of other investigators point to injury of the hypothalamus rather than of the posterior lobe as being chiefly responsible for the obesity. Marked obesity is also a feature of certain hypothalamic disorders in man unassociated apparently with any disease of the pituitary itself. Dystrophia adiposo-genitalis appears in two forms according to the age at which it develops—the *infantile* or *prepubertal* and the *adolescent* or *adult*.

THE INFANTILE OR PREPUBERTAL TYPE, FROHLICH'S SYNDROME (fig 57 15) This type may occur in children of any age before puberty. It may be the result of an inherent defect of the pituitary, of atrophy of the secretory cells by pressure (e.g., by tumors), injury (e.g., a penetrating wound) or of some infectious disease. Polyuria

and a high sugar tolerance are frequent accompaniments of the disease. The subjects are lethargic or somnolent and often of subnormal intelligence. They usually have voracious appetites and especially a craving for sweets. The "fat boy" of *Pickwick Papers* was undoubtedly an example of this condition. The younger the age of the child at which the disease commences the greater, obviously, will be the degree of stunting. When the dwarfing is of high grade this combined with the obesity makes a very striking picture. These subjects are human counterparts of Smith's rats in which the pituitary was destroyed (and the hypothalamus presumably injured) by chronic acid injections.

THE ADOLESCENT OR ADULT TYPE Male subjects of this condition are often effeminate in disposition and appearance. The excess fat has a feminine distribution, the adiposity being noticeable chiefly in the mammary region, buttocks, thighs and over the mons veneris. The hair over the pubis and in the axillae is sparse or absent, the skin of the face and the body is smooth, soft

and hairless, the hips are broad. In female subjects the obesity is often extreme, a weight of 300 pounds being not very unusual (fig 57 16). In both sexes the feet and hands are small and "pretty", the finger tips being slender and tapering with narrow pointed terminal phalanges. The extremities thus give a picture the reverse of that seen in acromegaly. The basal metabolic rate is often subnormal and sugar tolerance increased. Diabetes insipidus is a common, and narcolepsy (p 1028) an occasional, accompaniment.

LAURENCE-BIEDL-MOON SYNDROME

This condition, hitherto attributed to pituitary deficiency, is now believed to be due to involvement of the hypothalamus. Since it bears a resemblance to the condition described in the last paragraph it is convenient to consider it here. The chief features of the disease are obesity, sexual infantilism, retinitis pigmentosa, polydactylism, mental deficiency and a familial tendency (fig 57 17).

THE THYROID GLAND

DEVELOPMENT, HISTOLOGY, BLOOD AND NERVE SUPPLY

Development, histology, blood and nerve supply

Very early in its evolutionary history the thyroid had a digestive function, a function lost long since but which its ontogeny recalls. The gland is developed from a single median outgrowth of hypoblast derived from the ventral wall of the primitive pharynx at the level of the 1st visceral cleft. This extends downwards. Its lower end bifurcates and enlarges to form the isthmus and lateral lobes of the thyroid. Its upper end gives rise to the foramen cecum of the tongue. The intervening portion—the *thyro glossal duct*—normally disappears but sometimes persists and may give origin to accessory thyroids or to the so-called thyro glossal cysts.

The *thyroid tissue* is composed of cuboidal epithelial cells arranged in a single layer around spaces roughly circular or ovoid in shape and fairly uniform in size. These spaces, variously known as *follicles*, *vesicles*, *acini* or *alveoli*, contain a homogeneous gelatinous material—the *colloid substance*—which is the stored secretion of the gland. Connective tissue fibers support the alveolar walls and form septa which divide the gland into smaller masses. The cells lining the alveoli contain numerous mitochondria and a well defined Golgi apparatus. When the gland becomes active the Golgi apparatus hypertrophies and droplets of colloid appear in its proximity. Three types of follicle can be distinguished by special staining methods—*basophilic*, *acidophilic*, and *mixed*. In the first type the colloid is stained with basic dyes and appears blue. In the second type it stains pink with acid dyes. In the mixed variety, part of the colloid is basophilic and part acidophilic. These differences in the staining reactions of the follicles are dependent upon their degree of functional activity at the moment, the acidophilic type being the most active. The maximal normal weight of the human thyroid is, according to Marne, from 20 to 35 grams or around 0.4 gram per kilogram of body weight. The thyroid tissue of the early fetus consists of masses of epithelial cells showing little or no arrangement into acini. The latter appear about the 4th month but are small and contain little colloid.

The *blood supply* comes from the superior and inferior thyroid arteries, chiefly the former. The blood flow is profuse, the blood passing with little resistance from the arterial to the venous side through a wide capillary bed. The flow amounts to from 3.5 to 6 cc. per gram of tissue per minute, or about 5 liters per hour for the whole gland. The gland is richly supplied with *lymphatics* which drain lymph spaces surrounding the vesicles.

The *nerve supply* of the thyroid is derived from the vagus and the sympathetic. The sympathetic fibers leave the spinal cord between the 2nd and 5th thoracic segments and pass to cell stations in the superior and middle cervical ganglia, whence they are relayed to the gland through the superior laryngeal nerves and along the blood vessels. It is probable that the thyroid nerves are purely vasomotor in function and influence the activity of the gland indirectly, namely, by altering its blood supply. Control of thyroid activity is exerted mainly, if not exclusively, by the thyrotrophic hormone of the pituitary (p. 787).

AN OUTLINE OF THYROID FUNCTION

Some of the first experiments upon the thyroid were performed in 1856 by Schiff who found that in dogs death followed its removal, from the symptoms preceding death it is now apparent that this was due to removal of the parathyroids (ch. 60). As in the case of some other ductless glands, clinical observations gave the first hint concerning the functions of the thyroid. Hilton Fagg in 1871 reported a case of *cretinism* (p. 809) and ascribed it to absence or atrophy of the thyroid. Three years later Gull described the condition which today is known as *myxedema* or *Gull's disease*, but called it the *cretinoid state in adult life*. Horsley some years later removed the gland from monkeys and produced conditions resembling human cretinism and myxedema. Later experiments upon other species have established the fact that the thyroid secretion is absolutely necessary for the normal growth and development of young animals, and for maintaining the normal level of metabolism of animals of all ages. Magnus-Levy in 1895 demonstrated that thyroid deficiency was associated with a greatly reduced metabolism and that treatment with desiccated thyroid restored the metabolic rate to the normal level or above. Complete thyroidectomy reduces the basal metabolic rate by from 40 to 45 per cent in about 8 weeks after the operation. Thyroidectomy in lambs, young rabbits, goats or calves has been shown to cause retarded skeletal growth and arrested sexual development (fig. 58.1). Apathy, lack of vigor, thickening of the skin, and a striking reduction in basal metabolism result in both young and full grown animals. In young cattle increase in bulk but not in height is a notable

feature, the animals appearing short-legged and of a broad, stocky build. Other effects of thyroid removal in the young are, delay in the ossification of the epiphyses of the long bones, poor growth of hair, failure in thymic involution, and sub-normal intelligence. The administration of thyroid extract to young animals shortly after thyroidectomy prevents these otherwise inevitable results. In full-grown animals the effects of thyroid removal can be corrected by thyroid administration at any time after operation.

The effects produced upon lower orders by thyroid removal and by thyroid feeding are even more striking. Gudernatch removed the thyroids from tadpoles, keeping a number of young animals of



FIG 58 1 Triplet kids. Center animal normal, right- and left-hand animals thyroidectomized at 20 days old. Photograph taken 13 weeks after operation. (After Sutherland Simpson.)

This animal is allied to the frog, but is purely aquatic in its habits. It has, in its *adult* form, a finned tail, gills and four short limbs, resembling somewhat an enormous tadpole which has undergone partial metamorphosis. Thyroid feeding causes the axolotl to lose its fin and gills, develop air breathing organs and forsake the aquatic life for which Nature had designed it. See also chapter 57.

In growing birds (chicks) and mammals, a stimulating effect of the thyroid hormone on growth has been demonstrated. Acceleration of

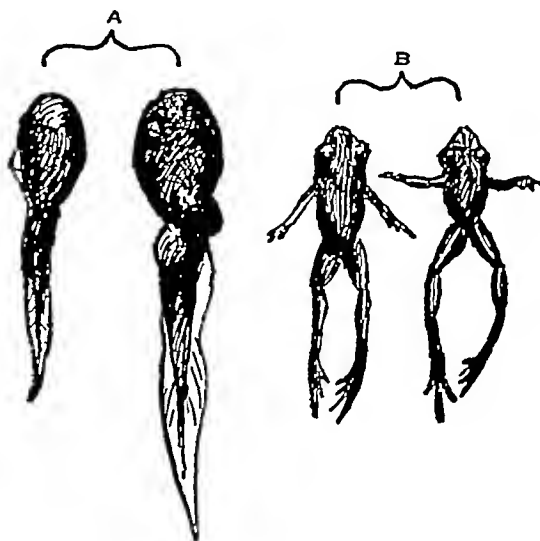


FIG 58 2 The effect of the removal of the thyroids upon the development of tadpoles. A, thyroidless tadpoles; B, normal frogs of the same age as A. (After Allen, redrawn.)

growth has also been observed in a twin child suffering from hyperthyroidism.

THE RELATION OF THE THYROID TO THE ADRENALS AND GONADS

The thyroid-pituitary relationship is considered in chapter 57. A relationship between the thyroid and adrenals is indicated by the following observations: (a) Subjects of hyperthyroidism show increased susceptibility to adrenal administration (Goetsch test). In animals, the threshold dose of adrenaline for cardiac acceleration is reduced by a previous administration of thyroid hormone. (b) Thyroxine (p. 821) administration to normal dogs causes hyperglycemia; this does not occur if the adrenal veins have been tied before injection of the thyroid hormone. (c) In hyperthyroidism, adrenaline instilled into the conjunctival sac causes

the same hatching as controls. The thyroidectomized larvae grew somewhat larger in size but did not metamorphose; the controls developed into frogs within the usual time (fig. 58 2). Metamorphosis of the thyroidless creatures could be induced to proceed at the normal rate by thyroid feeding.¹ Also, the time required for the normal larvae to metamorphose completely could be shortened (from 104 to 20 days) by feeding thyroid tissue. Swingle has shown that inorganic iodine alone will produce similar effects (p. 826). The effect of the thyroid hormone upon the axolotl is extraordinary.

¹ This effect has been made use of to assay the potency of thyroid preparations or to determine the activity of a particular thyroid gland (Gudernatch test).

dilatation of the pupil (Loewi's sign), an effect which does not occur in a normal person (d) Marine and Baumann found that injury to the adrenal cortex caused an increase of 60 per cent in heat production No change in metabolism occurred, however, if the metabolic rate had first been reduced as a result of thyroidectomy Marine suggests that the adrenal cortex normally exerts, through the pituitary, some inhibitory control over thyroid function

The following observations suggest an interrelationship between the thyroid and gonads (a) Thyroid enlargement is frequently observed at puberty and during menstruation or pregnancy (b) Castration in the dog or rabbit usually leads to a slow reduction in the size of the thyroid and a depression

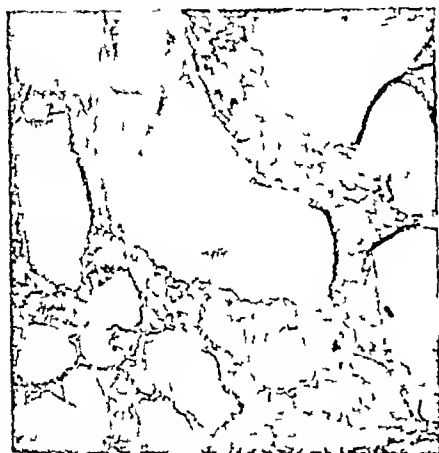


FIG 583 Photomicrograph of a human simple colloid goiter

of the metabolic rate (Marine) (c) The continued injection of estrogenic substances into rats or guinea pigs causes thyroid enlargement followed after a few days by involution of the gland (d) Thyroid feeding is said to inhibit estrus

These thyroid-gonadal relationships are probably brought about through the pituitary

GOITER

Goiter is a generic term which may be applied to almost any non-inflammatory and non-malignant enlargement of the thyroid gland The following is a short classification

- A. *Simple goiters* These are unaccompanied by constitutional features They are subdivided upon a histological basis into three groups
- (1) Colloid (diffuse)

- (2) Parenchymatous (diffuse)
 - (3) Adenomatous (nodular)
- B *Goiters associated with a deficiency of the thyroid hormone (hypothyroidism)*
- (1) Cretinism
 - (2) Myxedema
- C *Goiters associated with an excess of the thyroid hormone (hyperthyroidism)*
- (1) Exophthalmic goiter
 - (2) Toxic adenoma

Diffuse colloid goiter

The alveoli are large, distended with colloid and lined by low cuboidal or flattened epithelial cells There is no hypertrophy or hyperplasia of the latter (fig 58.3 and lower photograph, fig 58.8) The iodine content per gram of gland tissue is low but as a rule the total quantity in the enlarged gland is not far from normal Colloid goiter may become converted into, or result from, the following type (see also pp 816-819)

Diffuse parenchymatous goiter

Hypertrophy and multiplication of the cells lining the alveoli, with great reduction in the amount of colloid material are characteristic features of this type The alveolar cavities are of various sizes and shapes and often almost obliterated by infoldings of their walls The epithelial cells are high columnar instead of the normal cuboidal type The iodine content of the gland is low, usually much less than 0.1 per cent of its dried substance. Exhaustion of the gland and atrophy of its secretory elements with an increase in fibrous tissue may result, or, as mentioned above, the goiter may change to the colloid type, especially after iodine administration (fig 58.8) Partial thyroidectomy in animals, as first shown by Halsted, results in regeneration of the thyroid remnant to produce the foregoing histological picture of diffuse parenchymatous hypertrophy This observer also showed that removal of the thyroid from pregnant bitches led to parenchymatous goiter in the puppies

Adenomatous goiter

As a result of the formation of isolated tumor-like masses of thyroid tissue (adenomata) the glandular enlargement is asymmetrical or nodular The minute structure of the adenoma may resemble a section of colloid or of parenchymatous goiter, or it may undergo cystic changes Again, the alveoli may be unusually small, contain little

colloid and resemble fetal thyroid tissue (fetal adenoma) The iodine content of the nodule may be normal or high while that of the rest of the gland, which may also show diffuse colloid or parenchymatous changes, is usually low

The simple enlargement of the thyroid which sometimes occurs at puberty or during pregnancy may be either of the colloid or parenchymatous type A certain degree of thyroid enlargement at these times is physiological The enlargements seen in goiter districts (endemic goiter) or occurring sporadically may show the features of any of the

dry, thick, pasty and often deeply wrinkled, the nose is broad and its bridge depressed, the tongue is enlarged and appears between the thickened and usually parted lips, (5) the supraclavicular fossae are filled with pads of fat, (6) closure of the anterior fontanelle which normally occurs before the 20th month is postponed for several years, the epiphyses of the long bones fail to ossify at the usual time, (7) the basal metabolic rate is depressed by from 20 to 40 per cent below the normal

Endemic cretinism is much less common today In the past it was seen most frequently in districts

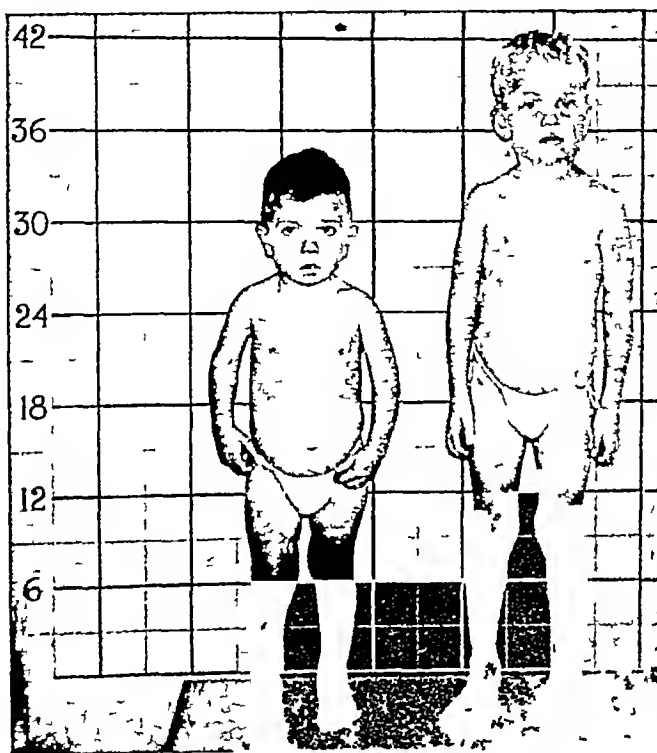


FIG 58 4 A case of cretinism of mild degree as compared with those shown in figure 58 5 (By the kindness of D A L Chute)

three types Either of the first two forms may end in exhaustion atrophy (p 816)

HYPOTHYROID STATES

CRETINISM

A cretin is a type of dwarf for which deficiency or absence of the thyroid secretion in infancy or early childhood is responsible (figs 58 4 and 58 5) Among the typical features of the condition are (1) retarded and abnormal skeletal growth, (2) arrested sexual development, (3) mental deficiency varying in degree but often amounting to complete idiocy, deaf mutism is common, (4) the facial features are coarse and appear bloated, the skin is

where goiter was prevalent—in the valleys of the Alps, Pyrenees, Himalayas, etc.—and was usually the result of atrophy and degeneration of the secretory epithelium of a goitrous gland Some of the worst cases of cretinism appear, however, in infants, either in these districts or elsewhere, who are not goitrous The condition is then due to (a) prenatal or early postnatal atrophy of the thyroid, (b) to its congenital absence, or (c) to its destruction by some inflammatory condition A large percentage of the goiterless cretins in a goiter district are the result of hypothyroidism in one or both of the parents

Though the majority of cretins are apathetic and

sluggish, a very few are highly excitable—the nervous cretinism of McCarrison

MYXEDEMA (GULL'S DISEASE)

Myxedema is the result of thyroid deficiency in adults or older children and corresponds to the cretinism of infants and younger children. It fol-

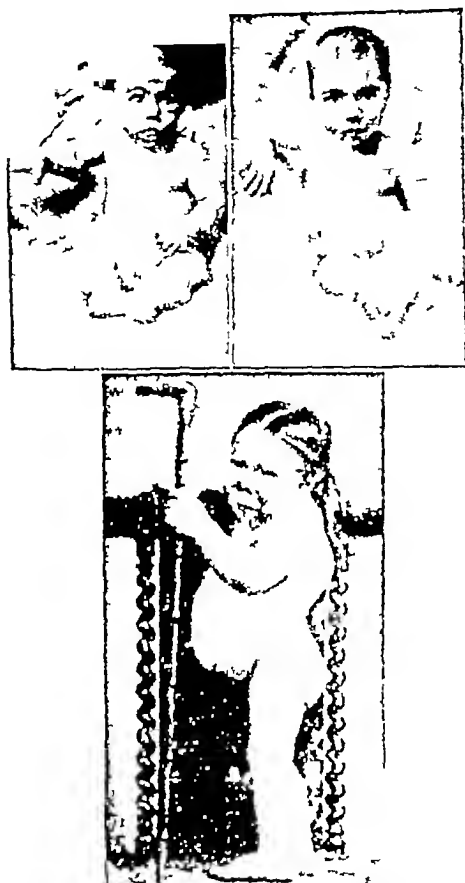


FIG 58.5 Upper photograph (left), cretinism in an infant (right), same child after treatment with thyroid extract. (After J Huxley) Lower photograph, a cretin aged 10 years (From McCarrison after Thomson)

lowers atrophy or destruction of the thyroid from whatever cause, or may result from an operation in which too much of the gland has been removed (*operative myxedema* or *cachexia strumipriva*). The chief characteristics are (a) low metabolic rate (-20 to -40), (b) a thick puffy appearance of the skin, the hair is dry, brittle and sparse, the facial features sometimes give the subject a mongoloid look (fig 58 6), (c) apathy, lethargy, slow

cerebration, though general intelligence is retained, (d) increased body fat and hypercholesterolemia, (e) greater susceptibility to cold, and (f) cardiac dilatation and low voltage electrocardiograms. The edema-like appearance of the skin was thought by Ord, who suggested the term myxedema, to be due to the accumulation of mucin in the subcutaneous tissues. The thickening of the skin is not due to a true edema, nor is it the result of the accumulation of mucin but, according to Boothby, to the deposit of a semi-fluid albuminous substance containing over 13 per cent of protein, or about the concentration of protein in white of egg. This deposit is looked upon by Boothby and his associates as representing an increase in the normal quantity of stored or reserve protein (p 632).

A non-myxedematous condition associated with a 20 per cent reduction of the basal metabolic rate, undue susceptibility to fatigue, increased sensitivity to cold and sometimes nutritive abnormalities of the skin, hair and nails is recognized. It is ascribed to a relatively mild grade of hypothyroidism.

HYPERTHYROID STATES, THYROTOXICOSIS

EXOPTHALMIC GOITER (GRAVE'S, PARRY'S OR BASEDOW'S DISEASE)

In Graves' disease there is nothing in the histological appearance of the thyroid to distinguish it, with certainty, from a simple goiter. The gland usually shows a picture typical of parenchymatous goiter, i.e., hypertrophy and hyperplasia (p 808), its iodine content is low. The blood iodine is elevated. After iodine administration the gland tends to assume the histological appearance of colloid goiter (fig 58 8 and p 808). The blood supply of the gland is greatly increased, the rush of blood through the superior thyroid arteries often producing a loud bruit or a distinct thrill. In addition to the thyroid enlargement the chief features of the fully developed condition are

(a) Accelerated pulse (100 to 160 per min) and increased circulation rate (50 per cent or more above normal), cardiac dilatation and hypertrophy, myocardial failure, auricular fibrillation in 20 per cent of cases, flushing of the skin, normal or low diastolic pressure with high systolic pressure (i.e., high pulse pressure)

(b) Nervous excitability

(c) A fine involuntary tremor

(d) General muscular weakness

(e) Protrusion of the eyeballs (exophthalmos) and other ocular signs, e.g., widening of the palpebral

lowers atrophy or destruction of the thyroid from whatever cause, or may result from an operation in which too much of the gland has been removed (*operative myxedema* or *cachexia strumipriva*). The chief characteristics are (a) low metabolic rate (-20 to -40), (b) a thick puffy appearance of the skin, the hair is dry, brittle and sparse, the facial features sometimes give the subject a mongoloid look (fig 58 6), (c) apathy, lethargy, slow

fissure (Dalrymple's sign) due to retraction of the upper lid, infrequent winking (Stellwag's sign), weakness in convergence of the eyes (Moebius), tremor of the closed lids (Rosenbach), lid lag (von Graefe) i.e., failure

thyroxine (p 821) in excess, beats at the increased rate when perfused, or transplanted to the body of another animal. Also, fragments of heart muscle of a two-day-old chick embryo pulsate at a more



FIG 58.6 Left, myxedema. Right, same subject after three months treatment with thyroid extract (After Murray.)

of the lid to follow the eye smoothly upon looking down, absence of wrinkling of the forehead when eyes are rotated upwards (Joffroy), difficulty in eversion of upper lid (Gifford). There is frequently edema of the conjunctiva (chemosis) and upper lid, and palsies of ocular muscles may occur (fig 58.7).

(f) Metabolic rate increased to varying degrees up to 80 per cent above normal. As a result of the increased activity of the heat-dissipating mechanisms the skin is hot and moist. Increased tolerance to cold and lowered tolerance to a high environmental temperature.

(g) Dissipation of the fat stores, wasting.

(h) Nitrogen and calcium excretion are increased (rarefaction of the skeleton may be demonstrated by X-rays in many cases).

(i) Disturbance of carbohydrate metabolism is common, as evidenced by hyperglycaemia, glycosuria and reduced sugar tolerance. Hepatic glycogen stores are reduced.

(j) The thymus is enlarged in 80 per cent of cases.

(k) Subjects of hyperthyroidism are especially susceptible to oxygen deficiency. Work is performed less economically, and dyspnoea occurs upon exertion (p 418).

A DISCUSSION OF THE PATHOGENESIS OF SOME OF THE FEATURES LISTED ABOVE

The tachycardia is not dependent upon nervous connections but appears to be due to a direct and persistent effect of the hormone upon the cardiac muscle. The excised heart of an animal given



FIG 58.7 Typical case of exophthalmic goiter showing characteristic facies (After Crile.)

rapid rate when thyroxine is added to the nutrient fluid. The high pulse pressure seen in hyperthyroidism is due to the general vasodilatation combined with an increased stroke volume. The increased circulation rate is chiefly the result of the higher metabolic rate.

Cardiac failure and thyrotoxic disease. The association of cardiac failure with hyperthyroidism

has lead to the idea that the thyrotoxicosis is a cause of myocardial disease. But organic disease of the heart is not as common in Graves' disease as has been generally supposed. When cardiac failure does occur it is in the great majority of instances due to disease of the heart which had existed, though often without symptoms, before the development of thyrotoxicosis. The greater cardiac work (increased cardiac output) resulting from the higher metabolism, it is believed, by encroaching on cardiac reserve, merely reveals a hidden cardiac condition or aggravates a pre-existing manifest heart lesion. Heart failure in the presence of thyrotoxicosis, therefore, appears at an earlier age than it otherwise would. This is the view held by such authorities in the field as Hurvall, Andrus and Means. Mayer and Sittler express this view succinctly when they speak of the effect of thyrotoxicosis upon heart disease as comparable "to that of a catalytic agent." Though auricular fibrillation is a common effect of thyrotoxicosis, it alone according to Means is of little clinical importance.

Exophthalmos, exophthalmic ophthalmoplegia
The pathogenesis of exophthalmos in Graves' disease has been a puzzling question for years. And though a final answer cannot be given, the possible factors concerned must be discussed. It will be necessary first to define exophthalmos. It is the forward displacement of the eyeball, and should not be confused with the staring or frightened expression caused by retraction of the upper lid. The latter may exist without exophthalmos. In order to ascertain whether or not exophthalmos exists and, if it does, to determine its degree, the distance at which the most forward point on the cornea lies anterior to the outer bony margin (external canthus) of the orbit is measured. Special instruments (e.g., Hertel's exophthalmometer) have been designed for this purpose. The distance in normal persons ranges from 14 to 16 mm., whereas in Graves' disease it may be over 25 mm.

One of the earliest theories of the cause of exophthalmos was based upon the contractile power of *Mueller's orbital muscle* situated at the back of the orbit or of the smooth muscle in the *fascia bulbi*. The tone of either or both of these groups of smooth muscle was thought to be increased as a result of the thyroid hormone having sensitized them to the action of circulating adrenaline. These structures are innervated by the sympathetic and consequently might well be subject to adrenaline action enhanced by the presence of excess thyroid

hormone. But though Mueller's muscle in the dog is well developed, and when stimulated through the sympathetic, causes a sharp forward displacement of the eyeball, the same muscle in man is a mere vestige, and quite incapable of contraction powerful enough to cause such an effect. Furthermore, true exophthalmos does not disappear after death as it surely would do were it solely due to increased muscular tone.

There is sufficient evidence to state quite confidently that *exophthalmos is not an effect of the thyroid hormone itself* and bears no relation to the signs and symptoms of thyrotoxicosis. It may occur in a most severe form after thyroidectomy, and thyrotoxicosis may be pronounced in its absence. Exophthalmos cannot be produced in animals by large and repeated doses of thyroid hormone, nor can it be induced in man by this means.² But this sign is readily caused in normal animals (guinea pig-) or thyroidectomized animals, and even to a greater degree in the latter, by injections of thyrotrophin (p. 787). Furthermore, in patients in which exophthalmos persists after thyroidectomy, administration of the thyroid hormone does not aggravate it, indeed, there is evidence that the hormone induces an amelioration.

It is quite evident from these observations that the thyrotrophic hormone of the hypophysis, or an unknown principle closely associated with it,³ is the essential cause of exophthalmos in the guinea pig.

The mechanism causing the actual proptosis or forward displacement of the eye must now be considered. Smelser studied the minute appearance of the extra-ocular tissues in normal and thyroidectomized animals after the production of exophthalmos induced by injections of the thyrotrophic hormone. In both sets of animals he found increased bulk of the extrinsic ocular muscles and retro-ocular fat, due mainly to the accumulation of fluid. The dorsal lacrimal gland (Harderian gland), which is not present in man, was enlarged. Naefziger and Jones had previously described swell-

² Prolonged treatment with the thyroid hormone, as for obesity, may set up a thyrotoxic state, and exophthalmos sometimes develops as part of the syndrome. The exophthalmos, however, is not due to a direct effect of the administered hormone, but has the same pathogenesis as that of spontaneously developed Grave's disease.

³ According to Albert the pituitary contains an exophthalmogenic principle distinct from thyrotrophin but which runs parallel with it in activity. This factor produces exophthalmos in the minnow by the secretion of fluid into the retro-ocular space.

ing of the extrinsic ocular muscles in clinical exophthalmos Pochin, in a postmortem study of a number of human subjects of exophthalmos, found increased volume of the orbital tissues, but this was due largely to an increase in fat, both of the muscles and connective tissues The fat content of the muscles was increased by 50 per cent The accumulations of orbital fat were the more noteworthy because there was extreme wasting of the body generally Enlargement of the lachrymal gland contributed to the increased volume of the orbital structures These several observations make it quite clear that exophthalmos is caused by a rise in intraorbital pressure due in turn to the greater bulk of the retro-ocular tissues

From these and other similar observations it seems well established that increased bulk of the orbital contents and the resulting rise in retro-orbital pressure, is the immediate cause of both experimental and clinical exophthalmos In the clinical condition, weakness of the ocular muscles which share in the general muscular asthenia of Graves' disease is probably a contributing factor But that the thyrotrophic hormone or an associated principle is the cause of the orbital changes in Graves' disease, as it is of exophthalmos in guinea pigs, has not been finally proved There are certain dissimilarities between the clinical and the experimental condition which give some cause for doubt or at least for caution Enlargement of the Harderian gland, for example, constitutes largely, though not entirely, the increase in the bulk of the orbital contents in the guinea pig, and though enlargement of the lacrimal gland does occur in human exophthalmos, it is not as a rule a prominent feature Nevertheless, the suspicion is very strong that a principle derived from the hypophysis is responsible for the exophthalmos in Graves' disease Suspicion has been strengthened by Rawson's finding that thyroid tissue from patients in whom exophthalmos was a prominent feature possessed much less power, or none, to inactivate thyrotrophin, whereas inactivation was shown to a conspicuous degree by thyroid tissue of normal persons or of patients with thyrotoxicosis, but without exophthalmos These facts go a long way to explain the appearance of exophthalmos, or its aggravation after thyroidectomy

It is obvious then that there is too much incriminating evidence to exonerate entirely the pituitary, the only possible verdict is the "not proven" of Scottish courts

Malignant exophthalmos is a very severe, and intractable form of exophthalmos, accompanied by ocular palsies Brain prefers to call this condition *exophthalmic ophthalmoplegia* The exophthalmos and weakness of the ocular muscles may be unilateral The palsy usually involves muscles moving the eyeball in a single plan, e.g., rotation upwards and downwards The signs and symptoms of thyrotoxicosis are often mild, or, as after thyroidectomy, absent, in some cases, this severe form of exophthalmos occurs without evidence at any time of thyroid involvement Proptosis of the eyeball is extreme, the antero-posterior distance from orbital margin to apex of the cornea may be 30 mm or more Drying of the corneal surface, ulceration or optic nerve injury is likely to result The muscular weakness has been thought to be due to a coincident myasthenia gravis (which is not uncommonly associated with thyrotoxicosis), but its failure to be influenced by prostigmine disproves this supposition Nor is it due to a nervous lesion for it does not conform to the distribution of any ocular nerve The weakness is attributed by Brain to the extreme stretching to which the muscles are subjected by the accumulation of fat in the cone-shaped space formed by the muscle bellies as they pass from the eyeball to their bony attachments

Retraction of the upper eyelid is produced in a manner quite different from that which causes exophthalmos, being often present when the latter is absent It is thought to be due either to increased tone of Mueller's *palpebral* muscle, which in conjunction with the levator palpebrae superioris elevates the upper lid, or to lessened opposition of striped muscle antagonists (orbicularis oculi) weakened by the action of excess thyroid hormone The increased tone of Mueller's palpebral muscle is attributed to the action of the thyroid hormone in sensitizing this smooth muscle to adrenaline Such action of the thyroid hormone upon sympathetically innervated structures is well known Another example is the dilatation of the pupil which occurs in hyperthyroid patients, but not in normal persons, when adrenaline is instilled into the conjunctival sac (Loewi's sign) The other lid signs are apparently due also to a disorder of those muscles which constitute the lid-closure mechanism and are thus probably due to excess thyroid hormone and not to the thyrotrophic principle

The *general muscular weakness* seen in thyrotoxicosis is of two types One type is an associated

but independent muscular dystrophy which has become aggravated by the thyrotoxicosis. In the other category the weakness of the muscles is the direct effect of the thyrotoxicosis, in this type there is, as a rule, little muscular wasting. In both forms there is a defect in creatine-creatinine metabolism (p 634), in the second, but not in the first, thyroidectomy, or the administration of iodine or of thiouracil abolishes the creatinuria.

Thyroid crises During the course of exophthalmic goiter intense exacerbation of the symptoms may occur, accompanied by nausea, vomiting, diarrhea, dehydration, high temperature, a great increase in heart rate, erythema, extreme nervousness, thrashing about in bed, muscular weakness, and sometimes delirium or coma.

Toxic adenoma

A simple adenomatous goiter (p 808) may undergo increased functional activity and produce the features of pure hyperthyroidism, i.e., those produced by the administration of thyroid extract in excessive amounts, or those described under exophthalmic goiter. In toxic adenoma the rest of the gland is usually atrophic as a result, probably, of the depressing effect upon it of the excess of circulating hormone.

DISCUSSION OF HYPERTHYROID STATES

Today, many, perhaps most, authorities on thyroid disease do not consider that hyperthyroid states can be placed in two categories based upon the pathological processes in the gland—diffuse hyperplasia and toxic adenoma. Exophthalmos may appear in toxic adenoma and it may be absent though the gland is diffusely hyperplastic and accompanied by signs of thyrotoxicosis. In either case exophthalmos may overshadow the thyrotoxic manifestations or vice versa. Moreover, the predominance of one or other of these two features may represent only a phase of the disease and, in any single case, one may subside and the other become the more outstanding. Means gives the designation, Graves' disease, to all hyperthyroid states and refers to that type or phase in which exophthalmos constitutes the chief manifestation, and the thyrotoxic manifestations are absent or mild, as *hyperophthalmopathic Graves' disease*.

In the past, many found it difficult to believe that there were not two separate hyperthyroid states, distinct in their pathogeneses, depending upon the histopathology of the gland. Toxic adenoma was thought to

cause pure hyperthyroidism, i.e., its manifestations were believed to be only those which an excess of thyroid hormone would produce, just as a parathyroid adenoma causes hyperparathyroidism, or a tumor of pancreatic islet cells causes hyperinsulinism. Two theories were proposed to account for that type associated with diffuse thyroid hyperplasia, thyrotoxicosis, and exophthalmos. Plummer was the first to separate hyperthyroidism into two clinical types and to offer an explanation. He suggested that in toxic adenoma a normal thyroid hormone was produced in excess, whereas in the exophthalmic type an abnormal secretion was formed as a result of the incomplete saturation of the thyroxine molecule with iodine. This theory conformed nicely to the supposition held at that time that, whereas iodine administration benefited the latter type it was relatively ineffective in toxic adenoma. The "two-secretion" or dysthyroidism hypothesis of Plummer was criticized by Marine, who suggested that in the exophthalmic form of the disease some extra thyroid factor, e.g., the sympatho-adrenal system or the pituitary gland was involved. The evidence for the thyroid stimulating hormone (TSH) of the pituitary being responsible for the development of exophthalmos has been cited elsewhere (p 812). Marine pointed out the following observations which could not be reconciled with Plummer's theory, (a) the extract of thyroid tissue from an exophthalmic gland is no more toxic, usually less so, than that prepared from a normal gland, (b) the extract from a hyperplastic gland has a toxicity proportional to its iodine content, which is the reverse of that which one would expect to find were Plummer's theory correct, (c) diiodothyronine, whose molecule contains less iodine (2 atoms) is less active than thyroxine which contains 4 atoms of iodine, and thyronine which is iodine free is inactive, (d) Marine and Rosen induced exophthalmos in *thyroidectomized* guinea pigs by means of injections of anterior pituitary extracts, (e) clinically, exophthalmos is found in the absence of any manifestations attributable to the thyroid hormone, and even after complete thyroidectomy, (f) The features of the thyroid crises suggest a profound disturbance of involuntary nervous centers, furthermore, some of the most severe crises are postoperative, i.e., after a large part of the gland has been removed.

THE ROLE PLAYED BY IODINE IN THYROID FUNCTION

Iodine is an essential element of the thyroid hormone and its administration exerts a profound effect upon the thyroid tissue. Simple goiter is due to iodine deficiency. This may be (a) an *absolute deficiency* in drinking water and food, (b) a *relative deficiency*, the iodine of the food being sufficient for the elaboration of the amount of hormone required under the ordinary circumstances of life but

insufficient under more exacting conditions, e.g., puberty, pregnancy, excessive protein of the diet, etc. (c) It is also possible that certain microorganisms in the intestinal tract may reduce the quantity of iodine absorbed from the food, this would explain McCarrison's observations upon the relationship between the incidence of goiter in India and infected drinking water. It is more likely, however, that infected water, if it does sometimes

goiter, either simple or exophthalmic, it may be as low as 0.25 mg or less per gram. In colloid goiters the general level is higher. Marine considers all goitrous enlargements, as examples of "a compensatory work hypertrophy", brought about by iodine deficiency. The demands made upon the gland to produce its hormone without an adequate iodine supply results in hypertrophy and hyperplasia, or adenomatous growth, and in some in-

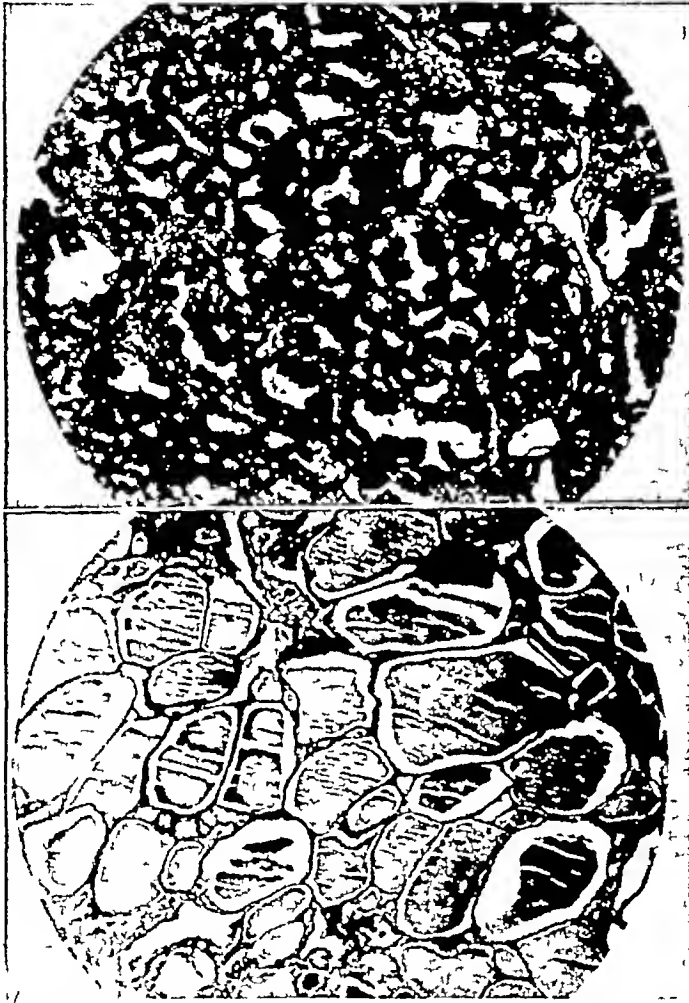


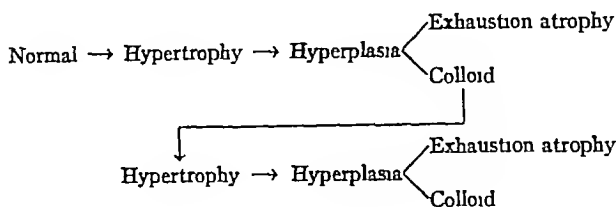
FIG 58.8 Upper photograph Microscopical appearance of the gland in exophthalmic goiter showing hypertrophy and hyperplasia (parenchymatous goiter) before the administration of iodine. Lower photograph, showing appearance after involution had occurred from iodine administration (colloid goiter) (After Rienhoff)

play a part in the production of goiter, acts in some other way than by merely interfering with the absorption of iodine.

The quantity of iodine in the normal human thyroid is about 2 mg per gram of dried tissue, the average total store in the gland is from 10 to 15 mg. A content below 1 mg per gram of dry gland, according to Marine, is indicative of definite thyroid abnormality, in severe parenchymatous

stances, final exhaustion and atrophy. This conception is well supported by the following facts: (a) The degree of hyperplasia of the gland is inversely related to its iodine content, but when hyperplasia gives place to exhaustion atrophy, the gland is almost iodine-free. (b) A colloid goiter, which is an expression of moderate iodine deficiency, passes into the hyperplastic type when the iodine intake is further reduced. (c) The paren-

chymatous hyperplastic picture is converted to the colloid type upon the administration of iodine (fig 58 8) Iodine thus induces the storage of colloid (thyroglobulin) The types of simple goiter are looked upon as stages in a process which is responsive to the iodine supplies, the gland sometimes passing several times through a cycle of hypertrophic, hyperplastic and colloid changes, or the cycle may end in atrophy The successive stages are represented in the following scheme after



Marine According to Rawson the thyroid may, in some instances, become restored to normal from the colloid goiterous state

The avidity of thyroid tissue for iodine, iodine metabolism The thyroid gland was shown some years ago by Marine to have a remarkable affinity for iodine Though constituting only about 0.05 or so of the body weight, the gland contains 20 per cent or so of the body's entire iodine content. The rest of the body's iodine store is contained in the skeletal muscles, mainly, but it is also in relatively high concentration in the pituitary and ovaries The entire blood contains less than a milligram Marine observed that when the thyroid was perfused with a solution of potassium iodide, relatively large quantities of iodine were taken up by the gland and could not be removed by subsequent washing This result was not obtained by the perfusion of other organs, e.g., spleen, kidney, etc The selective action of the intact thyroid for ingested iodine was also demonstrated Over 18 per cent of iodine fed to normal animals was recovered from the gland, the fetal gland also stores iodine fed to the mother

This earlier work has been confirmed and extended by the use of radio-active iodide (I^{127} , I^{130} and I^{131}) The last mentioned isotope which has the longest half life (8 days) is now employed almost exclusively After the injection of the isotope the thyroid gland if not already "saturated" becomes so within from 5 to 10 minutes

*"Saturation" is used in the physiological sense, namely the maximal amount of iodine which the gland will take up

The iodine is in the inorganic (iodide) form, but with the passage of time the iodide diminishes and the organic iodine concentration rises, indicating that the iodide has become bound, apparently to tyrosine groups (p 821) If a small dose (5 mg per kg) of radio-active iodide is injected into an animal whose diet is not particularly rich in iodine, from 50 to 80 per cent of the injected dose enters the gland The gland takes up about 80 times more iodine than does any other tissue Iodide appears

to be retained by the gland only in the ionized form If, instead of iodide, diiodotyrosine, thyroxine or the iodate is injected it does not enter the gland until some time has elapsed, apparently not until it has been broken down to iodide (Leblond and Sue) Hyperplastic glands show a greater than normal avidity for iodine The radio-iodine when injected into animals is found within an hour in the colloid of the thyroid follicles and to a less extent in the cells This iodine is all in the organic form—i.e., as diiodotyrosine or thyroxine In animals on an iodine deficient diet, the radio-iodine passes still more rapidly into the colloid, so rapidly indeed that little can be detected in the cells These observations are made possible by the use of the *radio autographic technic* Radio-iodine in contact with a photographic emulsion acts upon it to give a dark mark when the emulsion is spread upon a microscopic section of thyroid tissue which has taken up the isotope, the latter's position in relation to the histological structure can be observed (fig 58 9) The amount of iodide taken up by thyroid tissue is also increased by the thyroid stimulating hormone (TSH) of the pituitary but is reduced by certain antithyroid drugs (p 820)

Blood normally contains a total of from 5 to 10 micrograms of iodine per 100 cc. of serum, only minute traces are contained in the blood cells The iodine is present as iodide (inorganic) and in an organic form bound to the plasma proteins The inorganic iodine is only a small fraction of the total and since it varies with the iodine content of the

diet its determination is of less importance than that of the organic fraction. This latter consists of *di-iodotyrosine-like* and a *thyroxine-like* fraction. In extreme hypothyroidism the organic serum iodine

According to Salter, the level of the organic iodine of blood is a more reliable index of hyperthyroidism than the metabolic rate, for the latter is not controlled exclusively by the thyroid.

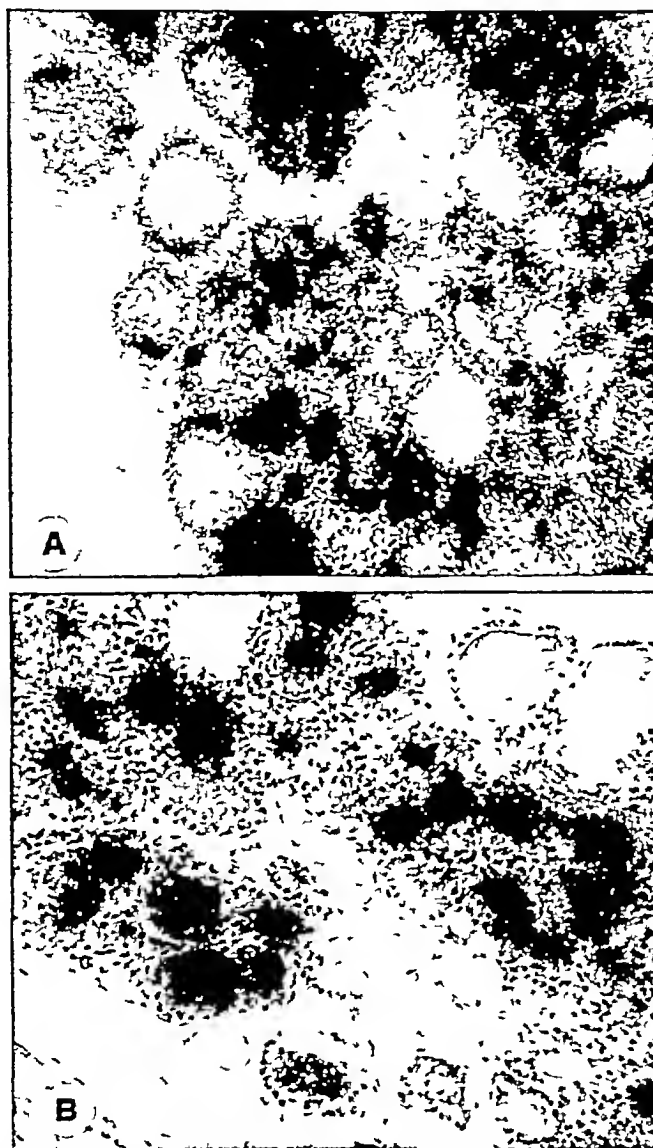


FIG 58.9 Radio-autographs of rat thyroids stained with hematoxylin and eosin. The location of radioactivity is indicated by the accumulations of black granules. In these sections the radioactivity is present as radioactive iodine incorporated in the thyroglobulin molecule.

A: Thyroid of a rat sacrificed one hour after injection of radio-iodide. In most of the follicles the radioactivity is arranged in the form of rings, because the newly-formed radioactive thyroglobulin is present in the epithelium of the thyroid follicle. B: Thyroid of a rat sacrificed 24 hours after injection of radio-iodide. The photographic reaction is present over the colloid of the thyroid follicles, indicating that the thyroglobulin formed in the cells at one hour is deposited in the colloid some 24 hours later. (After Leblond and Gross. Kindness of Dr. C. P. Leblond.)

is low, 1 to 2.5 micrograms per 100 cc, muscle iodine is also greatly reduced in cretinism or myxedema. In hyperthyroidism the organic iodine in serum ranges between 20 and 30 micrograms per cent. The iodine of fresh muscle is around 5 micrograms per cent, and is present in the same forms as in blood and in the thyroid gland.

After the administration of iodide, that which is not fixed by the thyroid gland is excreted in the urine, and into the stomach and upper intestinal tract. Little, however, appears in the feces, for most of that which finds its way into the stomach and upper intestine is reabsorbed through the

gastric mucosa and, to a greater extent from the lower intestinal tract

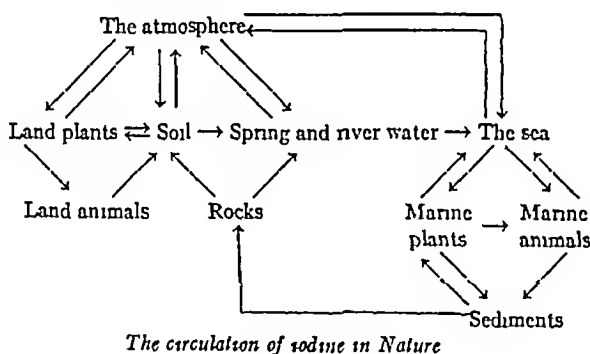
The daily iodine requirement of man is around 100 micrograms, but it has been shown that about 300 micrograms of thyroxine are used daily by the human body. This amount of thyroxine contains about 200 micrograms of iodine. In order then to explain the discrepancy between the daily iodine requirement and the amount utilized daily, it must be assumed that a part of the iodine liberated by the breakdown of thyroxine must be retained and utilized by the thyroid for the elaboration of fresh hormone.

Radio-iodine has recently come into use in the treatment of hyperthyroidism on account of its selectively destructive action on thyroid tissue.

Elmer has developed an *iodine tolerance test* based upon the especially high affinity of hyperplastic thyroid tissue for iodine. In performing the test as modified by Perkin and his colleagues 10 cc. of blood are withdrawn and its iodine content determined, 37 mg. of iodine in the form of Lugol's solution are then given

vogue. Coindet of Geneva in 1821 was the first to employ it in goiter. But as a result of its indiscriminate use, the true value of iodine was lost sight of. Chatin from 1850 to 1860 carried out some of the first scientific investigations into the relation ship between iodine and goiter, he showed that the iodine content of the soil, water and air of goitrous districts (valleys of the Rhone and the Seine) was very low. He attributed the thyroid enlargement to this deficiency, and recommended iodine administration as a preventive. Goiter is not seen along the sea-board. The sea contains an inexhaustible supply of iodine which has been leached from the soil. Sea water contains about 0.02 mg. of iodine per liter, fresh water, as a rule, very much less. The further away from the ocean and the more mountainous the country, the lower is the concentration of iodine in food and water and the higher in consequence is the incidence of goiter.

The accompanying schema modified from Lunde shows the distribution and circulation of iodine in Nature.



by mouth and the blood iodine determined in samples taken at half-hour intervals during the succeeding two and a half hours. The test, it is claimed, gives valuable aid in the diagnosis of borderline cases of thyroid disease. The curves drawn from the data are considerably higher for normal persons and for those with non-toxic adenomatous goiter, reaching a maximum of around 180 micrograms per cent, than for hyperthyroid subjects (max. 40 micrograms per cent).

Iodine in the prevention and treatment of goiter

Burnt seaweed or sponge, both of which are rich in iodine, have been employed in the treatment of goiter from the days of Hippocrates. Following the discovery of the element in the early part of the nineteenth century, its administration for the cure of many ills, but especially of goiter, became the

To Marine and his associates is due the credit for establishing the value of iodine in the prevention of goiter. They found that iodine or sea food prevented the thyroid enlargement in brook trout hatcheries. Marine and Kimball carried out experiments upon a large number of school children in Akron, Ohio, where goiter was endemic. They showed that of the group of children given iodine (2 grams sodium iodide in 10 daily doses twice a year) those who developed goiter amounted to only a small fraction of the number of goitrous individuals in a control group in which the iodine intake was not increased above that of the general population. The employment of small amounts of iodine in goiter districts such as Switzerland, New Zealand, Derbyshire in England and in parts of

the United States and Canada has proved to be a preventive measure of the utmost value. Many preparations of table salt have this essential added in a proportion of 1 part in 10,000 or so. By the use of this prophylactic, animal breeders in goiter districts have almost entirely eradicated the disease from their live-stock. Once goiter has become established iodine administration is of much less value but, as already mentioned, a hyperplastic parenchymatous goiter may be converted thereby to the less severe colloid type. The ingestion of iodine in much larger amounts than those mentioned above may induce hyperactivity in a simple goiter. Hyperthyroidism can be produced in animals by placing them upon a goitrogenic diet for a time and then administering iodide in excessive amounts.

Iodine in the form of Lugol's solution⁵ (10 to 40 minims daily) is invaluable in the treatment of exophthalmic goiter, hyperplasia gives place to the picture of colloid goiter (fig. 588), the symptoms abate, there is a pronounced fall in the basal metabolic rate and the danger of a thyroid crisis is reduced, or, if a crisis has commenced, it may be ameliorated or checked. In the Mayo Clinic, the surgical mortality has been lowered from 3.5 to 0.7 per cent and pre-operative deaths from 2.5 to less than 0.5 per cent since iodine treatment has been instituted.

Theories concerning the action of iodine in Graves' disease. The beneficial action of iodine in Graves' disease has been a puzzling problem and more than one theory has been proposed to account for it. Such an action seems, at first sight, to be quite inconsistent with its well recognized effect in preventing thyroid hypoplasia and hypothyroidism as well as with its being an essential component of the thyroxine molecule.

The most likely explanation of the amelioration of thyrotoxicosis by iodine is that it suppresses hormone production, both indirectly through an effect upon thyrotrophin liberation, and directly by inhibiting synthesis by the gland. An examination of sections of thyroid tissue before and after iodine administration shows a marked reduction in the hyperplasia in the latter instance. Also, excess iodide decreases thyroxine production by surviving thyroid slices. Furthermore, it is possible that iodine also diminishes the activity of the proteolytic enzyme in the thyroid (p. 825) and thus

hinders the absorption of the hormone into the circulation.

A consideration of factors other than an absolute iodine deficiency in the development of goiter. Goitrogenic compounds

The work of McCarrison and others indicates that dietary factors, other than iodine deficiency, are also concerned in the production of goiter. Diets deficient in the fat-soluble vitamins and in vitamin C appear to be conducive to its development. The ingestion of excessive quantities of fat or of protein (especially of liver) also predisposes to it.

In 1928 Chesney and his associates observed that rabbits which were on a diet of cabbage became goitrous. Other members of the *Brassica* family, e.g., cauliflower, Brussels sprouts, etc., are also goitrogenic. Rape seed was also found by Kennedy and Purvis to cause goiter in man. Cases of goiter in man resulting from a diet containing a large proportion of cabbage have been reported. A characteristic of vegetables of the *Brassica* family is their relatively high content of cyanogen compounds. Taking this hint Marine and his associates gave various cyanides to rabbits and obtained a marked thyroid enlargement, methyl cyanide was especially effective. The thyroid hyperplasia was accompanied by exophthalmos.

Marine and his colleagues suggested that the goitrogenic effect of cyanides was dependent upon their property of depressing tissue oxidations, increased thyroid function being a compensatory reaction instituted to oppose this action. It was further suggested that exogenous or endogenous cyanides or defective powers of the body to detoxicate such compounds might be a primary cause of exophthalmic goiter and that the compensatory reaction of the thyroid was brought about through the pituitary or the hypothalamic centers. Potassium thiocyanate is a more recently discovered goitrogen. Its action on the thyroid was first discovered in patients under treatment with the drug for arterial hypertension. This compound is a normal metabolite and, it will be recalled, is excreted in the saliva. It appears more likely that cyanides (also cyanates and thiocyanates) exert their effect upon the thyroid either directly or through the pituitary, rather than by depressing tissue metabolism generally.

The possibility of infected drinking water, in some instances, playing a rôle in the development

⁵ Iodine, 1 gram, potassium iodide, 2 grams, water, 30 cc

of goiter has been mentioned (p 815) The goitrogenic action of cyanides and cyanates is not due to a lack of iodine in the diet (absolute iodine deficiency) but they increase the demand for iodine (relative iodine deficiency), their effects being corrected by the administration of iodine

Thiourea and *thiouracil* and a large number of chemically similar drugs have been discovered within recent years which depress thyroid hormone production and lower the metabolic rate Other compounds, e g, sulfonamide drugs, potassium thiocyanate and ammobenzene compounds, including para-aminobenzoic acid, also inhibit thyroxine production The Mackenzies first demonstrated the anti-thyroid action of sulfaguanidine in

seen in hyperthyroidism These compounds are therefore peculiar in that they induce thyroid hyperplasia—increase in height of the epithelial cells, infolding of the acinar walls, and loss of colloid—together with low metabolic rate and other evidence of hypothyroidism These agents do not interfere with the uptake of iodide by the gland, but prevent the oxidation of the iodide to iodine, and, as a consequence, the iodination of the tyrosine molecule (p 822) The mechanism whereby the antithyroid drugs cause these effects is not agreed upon unanimously According to one view the action is indirect, namely, the inhibition of some oxidative enzyme system, e g, peroxidase. Others (Pitt-Rivers) believe that these drugs, ow-



FIG 58 10 Showing the effect of thiourea on growth *Above*, 84-day-old cretin rat (50 gm) from a mother which had been treated with thiourea for the entire gestation period *Below*, normal untreated animal (160 gm) of the same age. (After Goldsmith and associates.)

1943 Since then, over 100 drugs possessing anti-thyroid activity to varying degrees have been discovered Thiouracil and certain of its derivatives have been widely used in the treatment of hyperthyroidism Others are too toxic to be suitable for clinical use. Among those derivatives other than thiourea and thiouracil which show antithyroid activity are *propylthiouracil*, *methyl thiouracil*, and *thiobarbital* The antithyroid activity of the two last mentioned drugs is about twice as great as that of thiouracil. These drugs when administered to young animals produce effects comparable in every way to those caused by thyroidectomy, they cause what may be termed a functional or chemical thyroidectomy (fig 58 10) But they produce a histological appearance in the gland similar to that

ing to their antioxidant properties exert a direct chemical action, preventing the oxidation of iodide, as well as the oxidative coupling of the di-iodotyrosine molecules They do not inhibit or alter in any way the usual action of the thyroid hormone (TH) upon the tissue cells, for an animal responds in the characteristic way to administration of thyroxine They do not cause thyroid hyperplasia after hypophysectomy, i e., in the absence of the thyroid stimulating hormone (TSH) Intact animals after treatment with thiouracil and related drugs have enlarged pituitaries which contain no TSH, which points to excessive functional activity of the pituitary and rapid discharge of the latter hormone. Furthermore, the production of TH by surviving thyroid slices treated with these agents in the

presence of TSH is inhibited. From these established facts it is very generally believed that the thyroid hyperplasia is due to the excessive output of TSH as a secondary or compensatory effect of the fall in the concentration of the thyroid hormone in the circulation. But the observations of Rawson and his associates point to another effect of the thiol compounds which is perhaps of greater importance in causing thyroid hyperplasia. The thyroids of rats given thiouracil for only 24 hours showed hyperplasia at the end of this period, yet no fall in the blood level of protein-bound iodine occurred. Also, active TSH and thiouracil were administered simultaneously to cockerels, and activated TSH but no thiouracil given to a second group, while a third group received inactivated TSH and no thiouracil. In the first group the height of the acinar cells was over 50 per cent greater than that of the second group, and nearly 150 per cent greater than that of the third. These results taken together with the very rapid response to thiouracil in rats have led Rawson and his colleagues to conclude that thiouracil augments the action of TSH in the circulation.

Thiocyanates exert their antithyroid action in a somewhat different manner. They inhibit the uptake and concentration of iodine by the thyroid, their effect can be nullified by the administration of iodine. The goitrogenic effect of the thiol compounds is not affected by the administration of iodine, except in relatively large (pharmacological) dosage.

Stanley and Astwood have employed an ingenious method for the estimation, in man, of the anti-thyroid activity of a number of drugs based upon their property of inhibiting the accumulation of iodide in the thyroid. Radio active iodide (I^{131}) was administered orally and counts made of the gamma radiation by means of a Geiger-Muller counter placed upon the skin over the thyroid.

Antigoitrogenic factors Substances of unknown composition present in certain plant foods are antigoitrogenic, e.g., oats, ordinary lawn grass, and fresh alfalfa. Even cabbage contains an antigoitrogenic principle which is usually overbalanced by the goitrogenic property of its cyanogen compounds. In none of the plants just mentioned does the antigoitrogenic effect depend upon the presence of iodine.

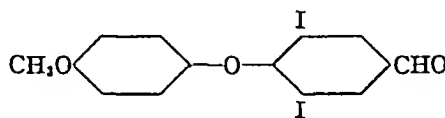
THE THYROID HORMONE

A glycerine extract of sheep's thyroids was first employed successfully by the English physician

George Murray in 1891 for the relief of myxedema. The thyroid tissue itself when given by mouth was shown subsequently to be physiologically active. Baumann in 1896 discovered that iodine was an important constituent of the thyroid extract. By acid hydrolysis of thyroid tissue and later by peptic digestion he obtained a brownish powder containing 10 per cent of iodine and possessing the physiological activity of the whole gland. He believed that the iodine was present in organic combination and named the compound *iodothyrim*. It was shown subsequently by Oswald (1899) that the active iodine constituent was attached to a protein—*thyroglobulin*—which is the chief component of the colloid material filling the alveoli of the gland. Thyroglobulin contains both diiodotyrosine and thyroxine joined to the protein molecule by peptide linkage. It is looked upon as the storage form of the thyroid hormone.

The active principle of the thyroid was isolated in crystalline form by Kendall in 1919. He named the substance *thyroxin*⁶ and found that it contained 65 per cent of iodine and an amino group. Harrington and Barger in 1927 established the chemical formula of thyroxine and effected its synthesis. It was found to be constituted of 2 benzene rings united by an oxygen bridge, and to contain 4 atoms of iodine and an amino-acid (alanine) side-chain. The synthesis required a number of separate steps.

3,5-di-iodo-4-(4'-methoxyphenoxy) benzaldehyde contains iodine atoms in the same positions as two of those in natural thyroxine—

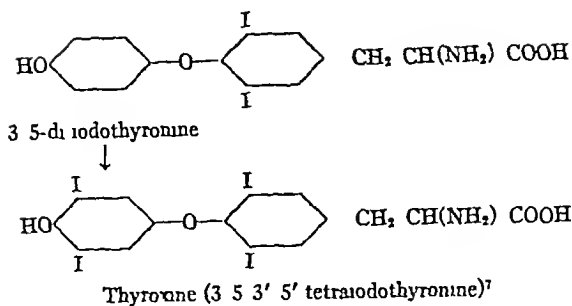


3,5-di-iodo-4-(4'-methoxyphenoxy) benzaldehyde

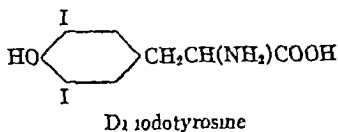
This compound having been prepared, the amino group was attached by condensation with hippuric acid. The resulting compound was boiled with NaOH in alcohol and then with acetic anhydride and hydriodic acid. The product of these procedures,—3,5-di-iodothyronine,—readily takes up 2 additional atoms of iodine with the formation of thyroxine, when treated with a concentrated solution of iodine in potassium iodide (see formula).

Synthetic thyroxine possesses physiological properties identical with those of the natural product.

⁶ Now usually spelled thyroxine.



Thyroxine is not formed in this manner in the thyroid gland, but from diiodotyrosine



This compound is formed in the body from tyrosine and certainly represents a stage in the synthesis of thyroxine, when added to slices of surviving thyroid tissue diiodotyrosine is converted to thyroxine. Diiodotyrosine itself has only slight physiological activity. Most of the organic iodine in thyroid tissue is in this form, a much smaller part being present as thyroxine. Four separate processes or steps are required for the synthesis of thyroxine, (a) the concentration of iodide in the gland, (b) the oxidation of iodide to elemental iodine, (c) the combination of iodine with tyrosine, and (d) the oxidative coupling of diiodotyrosine molecules, forming a diphenyl ether with the loss of one alanine side-chain. The thyrotrophic hormone (TSH) of the pituitary gland, appears to be necessary for all four steps. The mechanism responsible for the oxidative reactions, second and third steps, is not definitely known, but the most probable theory which has been proposed is that both the iodination of tyrosine and the coupling of two molecules of diiodotyrosine are effected through an enzyme system in which *peroxidase* plays an essential part. Fine granules in the thyroid cells which give the histochemical reaction characteristic of peroxidase have been described by Dempsey. The enzyme has not, so far, however,

been extracted from thyroid tissue, but in support of this theory is the fact that such antithyroid drugs as thiourea and thiouracil inhibit peroxidase action. An experiment performed by Keston is of great interest and suggests the type of reaction that may take place in the thyroid gland. When radio-active iodide was added with xanthine to milk and the mixture incubated, thyroxine containing the isotope was formed. The xanthine oxidase in the milk, acting presumably upon the xanthine, liberated hydrogen peroxide which in the presence of milk peroxidase oxidized the iodide to iodine which was introduced into tyrosine groups. Thyroxine was then formed by the coupling of diiodotyrosine molecules. If thiourea were added before incubation no thyroxine was formed.

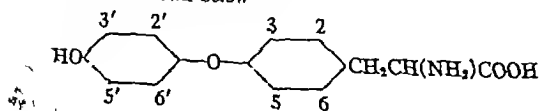
Additional evidence for the peroxidase theory is afforded by the observation of Harrington that hydrogen peroxide hastens the *in vitro* iodination of protein (see below).

Other enzyme systems, e.g., cytochrome, and tyrosinase, have been suggested as playing the chief rôle in thyroxine synthesis, but this is unlikely, for the cytochrome system is not inhibited by thiourea or thiouracil and tyrosinase has not been demonstrated in thyroid tissue nor does the inhibition of its action by thiourea derivatives run parallel with the antithyroid activity of these drugs.

The iodination of protein and the artificial synthesis of thyroxine

The remarkable discovery was made a few years ago that the iodination of casein, serum protein and other proteins, followed by hydrolysis, yielded crystals of thyroxine possessing the characteristic physiological activity. The first experiments in which the *in vitro* synthesis of thyroxine was demonstrated were performed by Abelson. His results were published in a series of papers from 1934 to 1938. His claim was not, however, generally accepted until the same result was obtained by

⁷ The numbering of the positions in the thyronine nucleus is shown below



Ludwig and Mutzenbecher in 1939 They gave an account of the method, a clear chemical description of the product obtained, and described its physiological property It was also shown that a small amount of thyroxine was formed spontaneously simply by prolonged incubation in alkaline solution These findings were soon confirmed by Harrington and Rivers and in several other laboratories Iodinated protein causes all the physiological effects of thyroid extract, relieves myxedema, and in large doses induces thyrotoxicosis

It has been thought unlikely until comparatively recently that thyroxine was identical with the hormone discharged from the thyroid into the blood stream It was argued that the very low solubility of thyroxine was against its being the true hormone It was also stated that the physiological activity of thyroglobulin or of the powdered gland was greater than could be accounted for by its thyroxine content

Harrington who at one time subscribed to this conception has since secured evidence by immunological methods that thyroxine itself is the natural thyroid hormone He and his associates succeeded in constructing an antigen in which thyroxine was combined as the active group (hapten) with protein In this thyroxine-protein complex the antigenic properties were borne entirely by thyroxine, the antigenicity of the original protein was completely lost Rabbits were immunized to this compound by a series of injections Serum from these animals was then given in a series of injections to normal rabbits When the latter were injected with a dose of thyroid hormone which is ordinarily effective, no rise in the basal metabolic rate was observed, though the same dose of hormone given to animals untreated with the antiserum caused a pronounced rise in heat production Thus, the thyroid hormone owed its specific activity to thyroxine It is probable that thyroxine rather than thyroglobulin is the form in which the hormone is present in the circulation Such a belief is strongly supported by the following consideration thyroglobulin is active when given by mouth, yet owing to the large size of the molecule it cannot be absorbed as such but must first be broken down in the intestinal tract Yet if it were the circulating hormone the unlikely assumption must be made that it undergoes resynthesis in the body in order to exert its activity Furthermore, Lerman, employing the delicate precipitin reaction has failed to demonstrate the presence of thyroglobulin in the circulation,

whereas, free thyroxine has been isolated from unhydrolyzed plasma (Taurog and associates)

Now, we know that the thyroid hormone is stored in the gland as thyroglobulin, and the belief that thyroxine is the circulating hormone leads to the question whether a proteolytic enzyme exists in thyroid tissue by which the thyroglobulin is broken down to permit the smaller thyroxine molecule to enter the circulation Such an enzyme has been demonstrated by De Robertis in the colloid of the gland, and was found to increase in activity by 100 per cent in thyrotoxicosis and to be reduced in hypothyroidism The thyrotrophic hor-

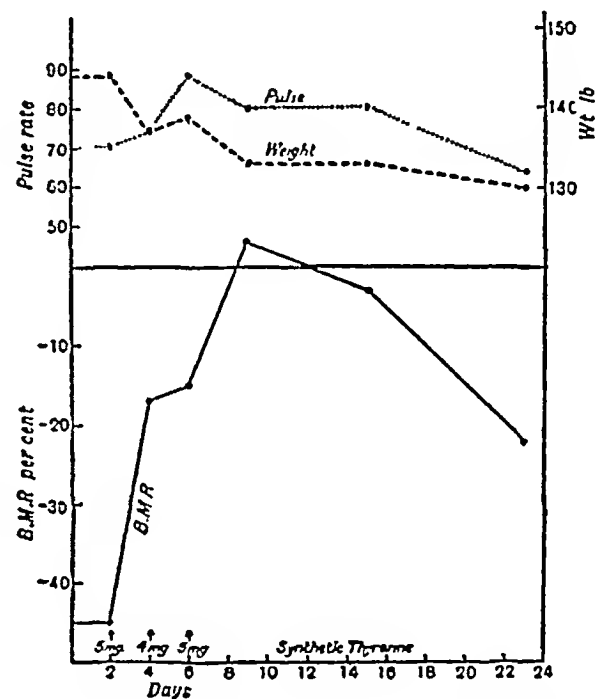


FIG 58 11 Effect of thyroxine on a myxedematous patient (After Harrington)

none of the pituitary activates this enzyme as the researches of De Robertis have demonstrated Within a few minutes after the injection of TSH, droplets appear in the cells lining the alveoli of the gland These intracellular droplets appear to be different in composition from the colloid of the alveoli and are believed to be smaller molecules derived from hydrolysis of the latter (see p 825)

The action of the thyroid hormone

The thyroid hormone is believed to act as a catalyst to increase the oxidative processes of the tissues This action most probably is brought about through its entering as an essential constituent

into some enzyme system. The oxygen consumption of tissue excised from a hypothyroid animal is subnormal, whereas the metabolism of tissue removed from a hyperthyroid animal is greater than normal. These observations indicate, of course, that the action of the thyroid hormone is exerted directly upon the cells rather than through the nervous system. The effect of a single administration of thyroxine upon the basal metabolic rate is slowly developed but prolonged (fig 58 11), it commences after the lapse of about 7 hours, reaches its maximum in from 8 to 10 days and lasts for 5 or 6 weeks. Thyroxine brings about the combustion of a relatively enormous quantity of material, 1 mg

of tissue cannot be explained by its property of increasing combustion in the cells, a substance such as dinitrophenol, which increases heat production, has no such action.

Animals when given repeated doses of dried thyroid, or of natural or synthetic thyroxine, develop the features of hyperthyroidism, e g, raised metabolic rate, wasting, increased excretion of nitrogen and of calcium (chiefly in the feces), hyperglycemia and glycosuria, reduction of liver glycogen, tachycardia, etc. The coincident administration of amino-acids enhances these effects. The administration of dried thyroid substance by mouth in the form of compressed tablets is a com-

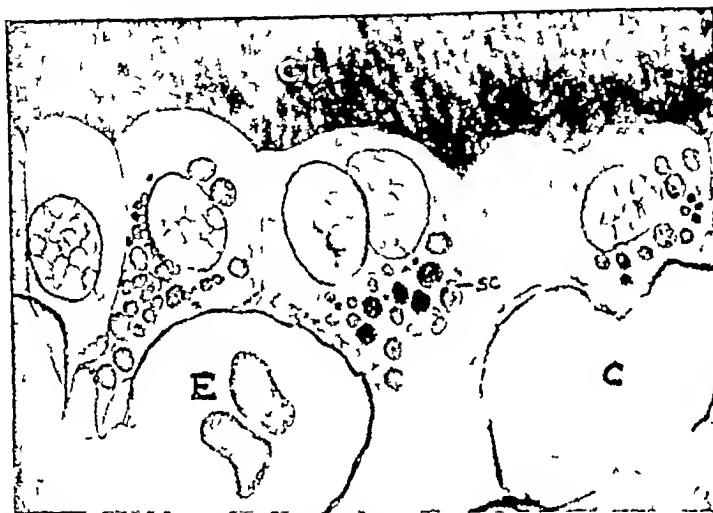


FIG 58 12 Cells of the thyroid of a rat which had been injected 22 hours previously with the thyroid stimulating hormone (TSH) of the pituitary. The intracellular droplets of colloid are increased in number and are collected toward the basal part of the cell from where they are discharged into the capillaries. The first effect of TSH, within 30 minutes or so, is to cause an increase in intracellular colloid which is accumulated toward that part of the cell abutting on the follicular lumen. The reversal of the direction of the movement of the droplets, i.e., from the free to the basal part of the cell commences in about six hours after the administration of TSH. *c*, colloid in lumen of the follicle, *c*, capillary, *E*, erythrocyte in capillary (After De Robertis)

causes the total CO_2 output to increase by some 400 grams and elevates the basal metabolism by 2.5 per cent. The entire normal adult human body contains about 14 mg of thyroxine. It has been calculated that the gland maintains this amount constantly by manufacturing and delivering about 0.33 mg per day. In other words, about 0.33 mg of the hormone is consumed daily, and as a matter of fact the daily quantity of administered thyroxine required to maintain the metabolism of a thyroidless adult at the normal level has been found to be about 0.5 mg. But the diversity of the effects of the thyroid hormone upon the development and differentiation of skeletal, nervous, and other tis-

plete corrective for myxedema. Cretinism also, if discovered early, can be immensely improved, and almost normal bodily and mental development induced (figs 58.5 and 58.6). Thyroxine is equally effective. The latter, in that it is a chemically pure substance, and so can be accurately measured, possesses an advantage over the dried gland, which is variable in potency.⁸ Nevertheless, thyroxine has

⁸ A given extract may be assayed either *chemically*, as by determining its content in iodine bound in thyroxine according to the method of Harington and Randall or *biologically*. Several biological tests have been used, e g, the rate of carbon dioxide production in mice, the rate of oxygen consumption of rats, or the increased sensitivity of the latter animals to oxygen deficiency, and finally the tadpole test of Gudernatch,

the greater disadvantage of being less potent and possessing a less certain action when given by mouth. This is due to its relative insolubility and its imperfect absorption from the gut.

THE SECRETION OF THE THYROID HORMONE

The thyroid hormone is formed in the cells lining the follicles and secreted as thyroglobulin into the follicular cavity where it is stored as the well-known colloid of the gland. It has been thought that the thyroglobulin was then passed, as required, into the blood along channels between the lining cells, or was even taken up again by the cells and transferred by them to the circulation. It was also held that at times of increased demand for hormone, thyroglobulin instead of passing into the follicle was secreted directly into the blood stream. As a result of improved methods of investi-

Though nervous impulses may possibly play some part in regulating the production and liberation of the thyroid hormone, it is now generally accepted that the chief and essential mode of control is by the thyrotrophic principle of the anterior lobe of the pituitary.

Factors in the regulation of thyrotrophin secretion

The factors regulating the production and release of TSH and the activity of the thyroid gland constitute a most complex system of which the details have not yet been fully disclosed. But there is good evidence for the following governing influences.

(1) Production or liberation of the thyrotrophic hormone is regulated by the level of the thyroid hormone in the circulation—a rise or fall depressing or stimulating, respectively, the production and

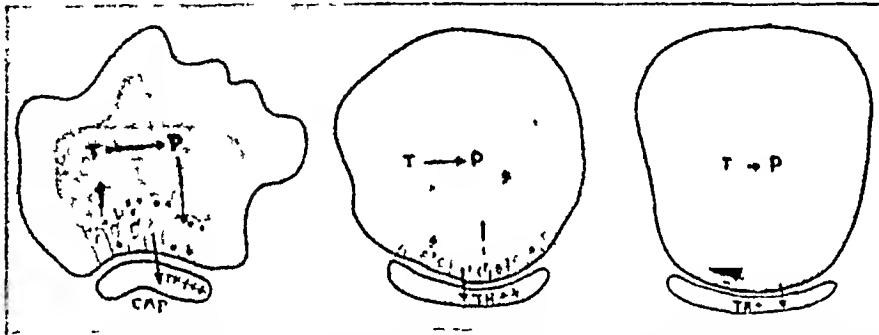


FIG. 58.13 Diagram of the enzymatic reabsorption of colloid under normal conditions (middle sketch), in a hyperactive follicle (left) and in a hypofunctional follicle (right). Length and direction of the arrows indicate the intensity of the processes of hydrolysis, and reabsorption, and of apical and basal secretion. T, thyroglobulin; P, product of the hydrolysis of T; TH, circulating thyroid hormone. The number of plus signs indicates the relative quantity of thyroid hormone. (After De Robertis.)

gation, especially by the use of radio active iodine, and the fact that thyroxine rather than thyroglobulin is thought to be the circulating hormone, this view finds few supporters today. It now seems clear that thyroglobulin represents merely a storage material and that once having entered the follicular cavity does not leave it again as such, but is hydrolysed by a proteolytic enzyme (see p. 823), the smaller molecule of thyroxine passing into and through the acinar cells into the blood (see figs 58.12 and 58.13).

in which the rate of metamorphosis is taken as the criterion. The last mentioned test is, however, dependent upon the total iodine content of the specimen (p. 807) and as we know, only a part of this is active for mammals. Axolotls have been used for a similar purpose. Thyroid feeding renders mice less susceptible to poisoning by acetonitrile (Reid Hunt). This fact has also been used as a basis for testing the potency of thyroid preparations.

liberation of TSH. It is uncertain whether the changes in the blood level of the thyroid hormone act directly upon the hypophysis, or indirectly through the hypothalamus. The selective retention of injected radioactive thyroxine by the pituitary has been demonstrated, which favors the first mode of action but does not, of course, exclude the possibility of an action via the hypothalamus.

(2) The thyrotrophic hormone is inactivated by the thyroid cell. Rawson found that TSH is inactive after exposure *in vitro* to normal thyroid tissue, but not when exposed to tissue of a simple goiter, tissue from a hyperplastic goiter of Graves' disease exerted the most pronounced inactivating effect. The tissue of lymph nodes and thymus had an effect similar to that of normal thyroid tissue. The observations on the inactivation of TSH by thyroid tissue accounts for differences in the

urinary excretion of TSH by normal, hyperthyroid and hypothyroid subjects. Normals excrete a small amount of the active hormone, hyperthyroid patients none at all, and those suffering from hypothyroidism due to goiterous thyroids, or as a result of total thyroidectomy, excrete large amounts. The

reverse of the effect on the thyroid of TSH or of the thyroid hormone on other tissues of the body.

(4) Iodide, in pharmacological dosage, inhibits thyroxine synthesis, an effect which is thought by Rawson to be due to inhibition of an oxidative enzyme system through which TSH enters the metabolism of the thyroid cell, or iodide may act as Robertis suggests by depressing the action of the proteolytic enzyme through which thyroxine is separated from the thyroglobulin molecule, and passed into the circulation.

(5) Adrenal corticoids (e.g., cortisone) or the adrenocorticotrophic hormone (ACTH) inhibits thyroid hormone production by depressing the action of TSH on the accumulation of iodine by the gland (Rawson). It is possible that they also depress the synthesis of TSH.

(6) Exposure of the body to cold causes hyperplasia of the thyroid and increased production of its hormone, an effect due to increased liberation of TSH brought about through the hypothalamus.

THE SYNTHESIS OF THYROXINE BY THYROIDLESS ANIMALS

The surprising discovery has been made that the thyroidless animal can synthesize thyroxine. Morton and his colleagues fed radioactive iodine to thyroidectomized rats and recovered radioactive thyroxine later from their bodies. This observation coupled with the fact already mentioned, that iodinated protein possesses thyroxine-like activity, has aroused the speculation that lower forms of animal life, which do not possess a thyroid, provided that they receive iodine in water or food, can synthesize thyroxine, or can utilize iodinated protein compounds as stimulants to metabolism. If it were possible to obtain a sufficient quantity of such compounds in the diet, higher animals also could probably dispense with the thyroid gland. The thyroid thus is seen as a highly specialized structure which more active forms of life have evolved for the conversion of compounds of low activity formed in the tissues generally, into a principle whose potency has been multiplied a hundred-fold over that of the original material.

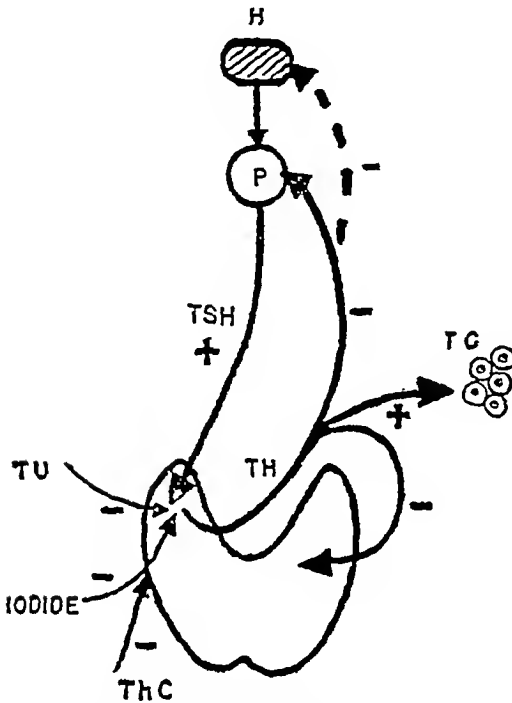


FIG 58.14 Diagram to show factors in the control of thyroid function. H, hypothalamus; P, anterior lobe of pituitary gland; TSH, thyroid stimulating hormone; TH, thyroid hormone; TC, tissue cells; TU, thyuracil; ThC, thiocyanate. Excitatory effects indicated by plus signs; inhibitory by minus signs. Broken line indicates possible effect of thyroid hormone on hypothalamus.

inactive hormone excreted in the urine of normal persons and in hyperthyroidism can be activated by means of reducing agents.

(3) The thyroid hormone also depresses some part of the mechanism—probably the action of TSH—through which it is synthesized. This effect, as shown by Galli-Mainini, is accompanied by a reduction in oxygen consumption, which is the

CHAPTER 59

THE ADRENAL GLANDS (SUPRARENAL CAPSULES)

DEVELOPMENT AND STRUCTURE

The mammalian adrenal gland, like the pituitary body and the thyroid-parathyroid apparatus, consists of two parts which, though closely associated anatomically, have separate origins, are structurally different and so far as is known, functionally independent. The central part of the gland is called the *medulla*, the outer enveloping rim of tissue is known as the *cortex*. In certain fishes (Elasmobranch) the analogues of these two parts are not joined together. Tissue corresponding to the mammalian medulla, for example, is found as a number of small discrete masses on either side of the spine in association with the sympathetic ganglia, while an elongated structure lying between the kidneys (*inter-renal body*) corresponds to the adrenal cortex of mammals. In the amphibia and reptilia the two types of tissue have come together, but masses of cortical cells are intermingled with islets of medullary tissue and the two tissues are not segregated into a peripheral and a central zone, as in the mammalian adrenal.

The medullary cells exhibit characteristic staining reactions, ferric chloride turns them blue, osmic acid black, and chromic acid or its salts a dark brown. As a result of the last mentioned reaction they are spoken of as chromophil or chromaffin cells.

In all animals the medullary tissue and the sympathetic ganglion cells have a common origin, they develop from primitive cell masses which have separated from the neural crest. Migrating from their sites of origin these masses of ectoderm cells undergo differentiation along two paths, some into sympathetic ganglion cells, others into chromaffin tissue. In the abdomen on either side of the mid-line a relatively large mass of chromaffin cells becomes enveloped by cortical tissue to constitute the adrenal medulla. Other smaller masses persist as accessory chromaffin tissue in association with the ganglia and plexuses¹ of the sympathetic. On the other hand, sympathetic ganglion cells may be found scattered among the cells of the adult adrenal medulla. The cortex is developed from mesoderm. It arises as a bud from the celomic epithelium covering the inner side of the fore part of the mesonephros. The celomic epithelium immediately behind this area gives rise to the germinal epithelium from which in turn the sex glands develop.

The medulla is composed of closely packed groups of polyhedral cells containing chromaffin granules which are looked upon as the mother substance of the medullary secretion. The cell groups are separated by blood sinuses which empty into a central vein. The cells of the cortex are arranged in three zones. These are from

without inwards the (1) *zona glomerulosa*, in which groups of cells are arranged in a circular or oval pattern, (2) *zona fasciculata*, in which the cells are arranged in columns, and (3) *zona reticularis*, which is composed of a network of cell cords (fig 59 1). The cells of the cortex contain fine droplets of *doubly refracting* lipid material which is also seen in the form of dust-like particles in the capillaries of this part of the gland, it is probably the active cortical hormone or its precursor.

The cells of the cortex originate at the periphery of the gland, in the *zona glomerulosa* just beneath the capsule, or more probably from the cells of the capsule itself. These young cells contain numerous mitochondria but relatively little lipid material. As the cells become older they migrate towards the medulla of the gland, the mitochondria become fewer and the lipid material more abundant. The lipid diminishes again later, the migrating cells become shrunken in appearance and finally die near the cortico-medullary junction. The lipid droplets contain cholesterol and insoluble ketones. The mitochondria are believed to be concerned in the production of the cortical hormone. Hypertrophy and hyperactivity of the adrenal cortex (induced by corticotrophin, p. 788, or by removal of one adrenal) is associated with proliferation of the mitochondria and a decrease in the lipid material. Adrenal atrophy (as follows hypophysectomy) is accompanied by a decrease in mitochondria and the lipid material appears to be increased due to its accumulation into larger droplets clumped together.

On the basis of lipid distribution, the normal adrenal cortex has been divided by Weaver and Nelson into four zones—an *outer zone*, comprising the *zona glomerulosa* and poor in lipid material, a narrow *optically inactive zone*, between the *zona glomerulosa* and the *zona fasciculata*, an *optically active zone* which is the richest in lipid material, its cells being well filled with fine dust-like particles, and an *innermost zone*, relatively poor in lipid and comprising the *zona reticularis* and a small part of the *zona fasciculata*.

A sex difference in the size and histological appearance of the adrenal cortex has been demonstrated in animals. The glands are larger in women than in men, a difference which is enhanced when the smaller female body is taken into account. In animals, the adrenals of the female are reduced to the size of the male glands under the influence of testosterone.

BLOOD AND NERVE SUPPLY

The adrenal is one of the most richly vascular organs in the body, receiving 6 to 7 cc. of blood per gram of tissue per minute. It is supplied by 3 small arteries

¹ Such chromaffin collections are called *paraganglia*.

which are derived, respectively, from the inferior phrenic artery, the renal artery and the aorta. These form rich plexuses in the cortex. The plexuses are continuous with the sinuses of the medulla which drain into the central vein of the latter. The right adrenal vein empties directly into the inferior vena cava, the left vein into the renal vein. The nerves are derived from the great splanchnic, the fibers pass through a plexus (suprarenal) before entering the gland. These fibers are medullated and have no cell stations in their course. That is, they are entirely preganglionic, the medullary cell itself taking the place of the ganglion cell and postganglionic fiber, they differ thus from all other sympathetic pathways (see p. 1102)

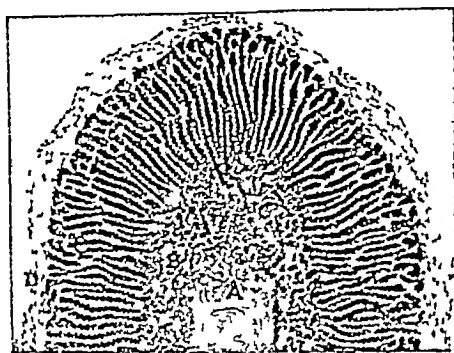
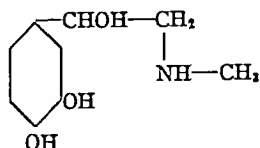


FIG. 591 Section of the human suprarenal capsule. A, medullary portion, B, zona reticularis, C, zona fasciculata, D, capsule, the zona glomerulosa is situated beneath the capsule (After Maximow and Bloom.)

THE ADRENAL MEDULLA

Thomas Addison's report in 1855 (p. 809) and the experimental work of Brown-Séquard in 1856 gave the first hints concerning adrenal function. The last-mentioned observer showed that complete removal of the glands from rabbits caused death. Not until many years later was it shown that the effects described by these pioneer investigators were due to loss of the adrenal cortex. The French physiologist Vulpian, working at about the same time as Addison and Brown-Séquard, discovered that the medulla, unlike any other tissue, was stained blue by ferric chloride, and that the blood of the adrenal vein sometimes gave a similar reaction. The staining reaction, it is now known, is characteristic of the internal secretion of the medulla (adrenaline). In 1894 Oliver and Schafer obtained an aqueous extract from the medulla which upon injection caused a pronounced rise in blood pressure.

The active principle of the extract was obtained in pure form in 1901 by Takamine and by Aldrich. Upon analysis the latter observer found the empirical formula to be $C_9H_{13}O_3N$. The substance has been given various names—adrenaline (or adrenalin), epinephrine, adrenin and suprarenin. The first of these names is most commonly used. Adrenaline is closely related to tyrosine, as will be seen from the following structural formula:



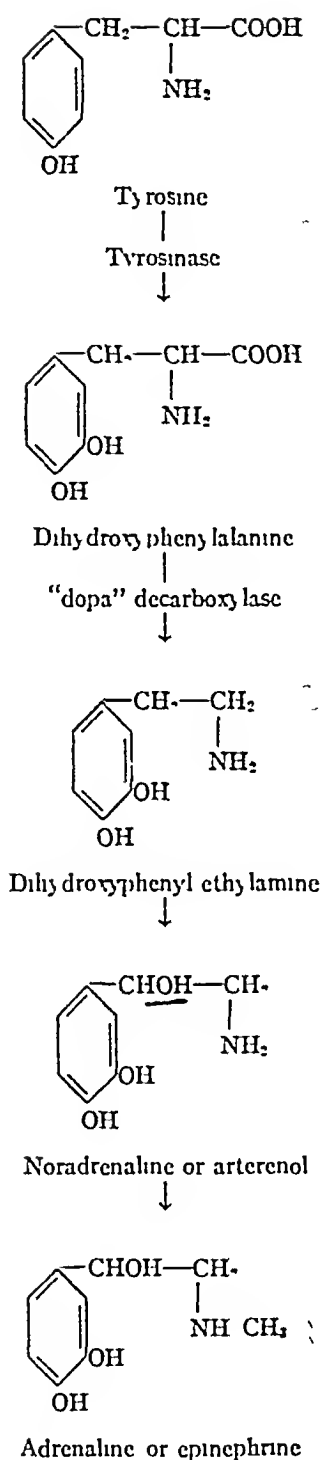
Adrenaline or epinephrine

Adrenaline is a secondary alcohol, its full chemical name being 3,4-dihydroxy- α -phenyl β -methylaminoethanol. It was first prepared synthetically by Stoltz (1904) and later by Dakin (1905). Adrenaline possesses an asymmetric carbon atom, so three isomers are possible (i.e., a levo and a dextro-rotary and a racemic form). The natural levo compound is 15 times more powerful than the dextro rotary form.

The intermediate steps in the synthesis of adrenaline by the adrenal medulla are not definitely known. But tyrosine is most probably a precursor and dihydroxyphenylalanine ("DOPA") produced by its oxidation is supposed to be one of the intermediate steps. The introduction of a second hydroxyl group into the tyrosine molecule can be effected and "DOPA" produced by the action of the enzyme tyrosinase, present in plants. The enzymatic decarboxylation of "DOPA" (by "DOPA" decarboxylase) could then form dihydroxyphenylethylamine (oxytyramine) and by the introduction of an hydroxyl group into this compound noradrenaline (arterenol or norepinephrine) could be formed. The addition of a methyl group to the latter would then yield adrenaline. This last step has been demonstrated by Bülbring who found that a suspension of adrenal tissue can convert, by methylation, noradrenaline to adrenaline.

The formulae on page 829 show the possible chemical steps in the production of adrenaline.

The belief that tyrosinase was confined to plants has been refuted in recent years by the demonstration of its presence in animal tissues as well. It has also been shown that DOPA can be produced from tyrosine by the action of ultraviolet light in the presence of Fe^{++} ions or ascorbic acid.



INACTIVATION OF ADRENALINE

Adrenaline disappears rapidly after injection, due to its oxidation in the tissues. The effects following subcutaneous injection are much less intense, though more prolonged, than those induced by intravenous administration. When given orally, adrenaline is inert. Adrenaline

is rapidly inactivated in the body. One site of inactivation is the liver, where presumably it is oxidized by an amino-oxidase. A smaller proportion is conjugated with sulphuric acid. Oxidative inactivation as by the cytochrome system occurs also and perhaps to an even greater extent in the tissues generally, which accounts for the evanescent action of adrenaline following injection. Outside the body adrenaline is much more stable in blood and tissue fluids than in Ringer solution, in which it undergoes oxidation to form a red compound. The substances in blood and tissue fluids which have a stabilizing or protective action upon the adrenaline molecule are reducing agents, such as glutathione and ascorbic acid (Welch). It is significant that the adrenal gland is especially rich in these substances.

ACTIONS OF AN EXTRACT OF THE ADRENAL MEDULLA, ADRENALINE AND NORADRENALINE

Not until nearly 50 years had passed since the pressor effect of the adrenal medulla was demonstrated by Oliver and Schafer, was it discovered that noradrenaline (arterenol, norepinephrine) as well as adrenaline was present in the gland. It has also been found by von Euler in postganglionic sympathetic nerves. Adrenaline and noradrenaline are present in the adult human adrenal medulla in the proportion of about 5 to 1, but noradrenaline is in higher proportion than this in the glands of young children, and in very large amounts (up to 90 per cent) in some adrenal tumors. It is probable as Blaschko suggests, and as is indicated by the formulae chemical, that noradrenaline is the immediate precursor of adrenaline. The discovery of noradrenaline has led to a revolution in our ideas of the actions of the medullary secretion but, like most revolutions, has been followed by a period of uncertainty and confusion. The two hormones, though closely allied in action, show certain important differences. For many years research has been carried out on adrenal medullary extracts in the belief that its only active principle was adrenaline. It will be necessary, therefore to describe the actions of the pharmacopeial or commercial preparations, for the most part, and where possible in the light of modern research to state which actions are due to adrenaline, and which to noradrenaline. At the outset the generalization may be made that with the exception of the effects on the vascular system, and on the pregnant uterus of the cat, the effects of noradrenaline, in so far as

they have been investigated, are weaker than those of adrenaline. Both agents act upon structures innervated by sympathetic nerve fibers and, with a few exceptions, the effects closely imitate those caused by stimulating sympathetic nerves. They do not act upon the nerve endings themselves, but on the effector organ, or perhaps some special substance in the effector (Langley). This hypothetical substance came to be known as the *receptive substance of Langley*. From this sympathetic-like action, adrenaline and noradrenaline (as well as a number of drugs) are said to be *sympathomimetic* in their effects.

(a) *Vascular effects* Ordinary preparations of adrenaline (e.g., USP epinephrine) have a pronounced pressor action causing widespread vasoconstriction. But adrenaline itself is by no means a general vasoconstrictor. Its overall effect upon

TABLE 79

Effects of adrenaline and noradrenaline on blood flow (ml/min)

After Barcroft, modified

	BEFORE ADRENALINE OR NORADRENALINE	DURING	
		ADREN	NORADREN
Liver	1500	3000	1500
Kidneys	1500	900	1200
Muscles	1000	2000	1000
Brain	750	900	675
Overall	4750	6800	4375

the vascular system is vasodilatation, that is, it decreases rather than increases the peripheral vascular resistance of the body as a whole. It constricts powerfully the arterioles and capillaries of the skin and the arterioles of the kidney, but dilates the vessels of the skeletal muscles and liver, and probably also of the coronary system. This vasodilator effect upon the muscles and liver overshadows the vasoconstrictor effects with a consequent decrease in the total peripheral resistance. The pronounced rise in blood pressure—a true adrenaline effect—is due entirely to its increasing the cardiac output. The rise in blood pressure is confined to the systolic phase, the diastolic phase showing no change or a fall—an expression either of the absence of any increase in, or a reduction of the total peripheral resistance (see table 79).

Noradrenaline is an overall vasoconstrictor,

causing an increase in the total peripheral resistance, and a hypertensive effect which is from 30 to 70 per cent greater than that of adrenaline. Noradrenaline constricts the cutaneous vessels, but to less degree than does adrenaline, the latter in man causing a much more pronounced pallor. Noradrenaline causes a reduction in blood flow through the brain of about 10 per cent, whereas adrenaline increases it by 20 per cent or so.

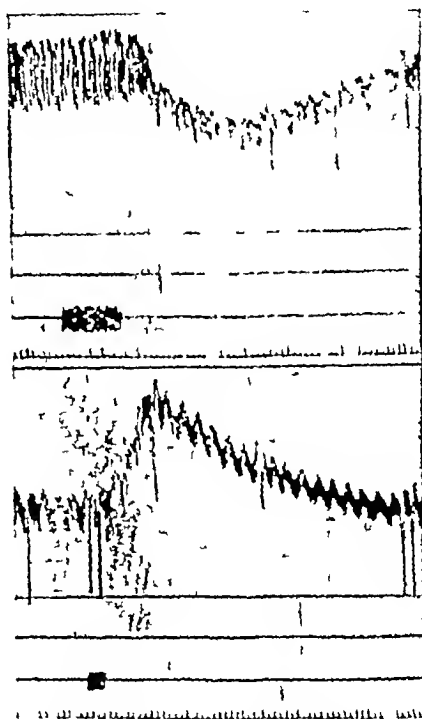


FIG 592 Lower tracing shows effect of intravenous injection of 0.02 mg adrenaline upon blood pressure. Upper tracing from the same animal shows the fall in pressure produced by the injection of 0.02 mg of adrenaline following a previous dose of 0.4 mg per kg of ergotamine tartrate. (After Geiling)

(b) *Heart* Adrenaline (USP epinephrine) exerts a direct stimulating effect upon the myocardium, it increases the oxygen consumption of the heart and augments the cardiac output mainly by an increase in stroke volume, this being responsible for the rise in systolic pressure. These effects are apparently due to the action of adrenaline as such, for noradrenaline either has no effect upon the cardiac output or decreases it. In animals with the vagus nerves intact, adrenaline slows the heart as a secondary effect of the hypertension (Marey's law) but causes cardiac acceleration after vagal

section In man, adrenaline accelerates the heart while noradrenaline causes bradycardia. Adrenaline (U S P epinephrine) speeds auriculo-ventricular conduction, the action of noradrenaline in this regard is unknown. Adrenaline also appears to increase the susceptibility of the ventricular muscle to the development of extrasystoles, or even of ventricular fibrillation.

(c) *Plain muscle* Adrenaline inhibits the muscle of the *stomach, intestine, bronchioles* and *wall of the urinary bladder*. Both the tone and movements of the intestine are inhibited (fig 59.3), the bronchioles are dilated. It excites the muscle of the *gall-bladder, urter, trigone* and *sphincter of the bladder*, the *retractor penis* and the *pyloric, ileocolic* and *internal anal sphincters*. The *uterus*, whether pregnant or non-pregnant, is contracted in man and in many animals, but in the cat, rat, mouse and guinea pig, the pregnant organ alone is contracted by adrenaline, the non-pregnant is relaxed. The contractions of the human uterus at term are often inhibited by adrenaline. Most of these effects are exhibited, though less strongly, by noradrenaline, with the exception of the excitation of the pregnant uterus of the cat upon which noradrenaline exerts a greater effect. As a result of the excitation of the radiating fibers of the *iris* (*dilator pupillae*) the pupil is dilated by adrenaline especially if the superior cervical ganglion has been previously excised. *Mueller's muscle* is stimulated and the eyeball protruded. Retraction of the *upper eyelid* is caused by the stimulation of its smooth muscle. The *nictitating membrane* of animals is retracted. Adrenaline also stimulates the *erectores pilae muscles* and other smooth muscle fibers in the skin.

(d) *Skeletal muscle* Adrenaline postpones muscle fatigue and increases the tension developed in a twitch of skeletal muscle. It increases muscle tone, and may cause tremors or muscular twitching. The effects of noradrenaline upon muscle are much less pronounced (see fig 65.3, p 964).

(e) *Respiration* After a short initial period of apnea the respirations are increased in rate and depth. The apneic period is apparently secondary to the rise in blood pressure and is brought about through the carotid sinus mechanism (p 283).

(f) *Carbohydrate metabolism* Adrenaline administered by injection causes hyperglycemia and glycosuria. It shows an antagonism to insulin—relieving hypoglycemic convulsions. These effects,

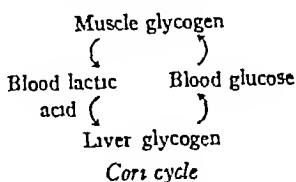
which are also shown by noradrenaline, but to a less degree, are due to the mobilization of sugar from the liver whose glycogen stores are thus reduced. The hyperglycemic effect is therefore greatest in well-fed animals with an abundant hepatic store of carbohydrate, and of course is not obtained in hepatectomized animals. The glycogen of the muscles is also reduced by adrenaline. On the other hand, there is evidence (Himsworth and Scott) that the rate of removal of glucose from the blood by the tissues is accelerated. When administered to animals after a prolonged fast or in other con-



FIG 59.3 Rabbit's intestine in Ringer's solution. At X Ringer's solution + adrenaline 1:100,000,000, at X' Ringer's solution substituted. Time 30 sec. (After Hoskins)

ditions which deplete the hepatic glycogen stores, adrenaline causes an *increase* in liver glycogen. The latter is due to the breakdown of muscle glycogen to lactic acid which, diffusing into the blood, is carried to the liver where resynthesis to glycogen occurs. There is a fall in the inorganic phosphate of the blood due apparently to the phosphorylation of glycogen and the formation of glucose monophosphate. If the adrenaline administration is continued the liver glycogen is converted in turn to glucose which passes into the blood, causing hyperglycemia, it is subsequently reconverted to glyco-

gen in the muscles (Cori) The cycle may be represented in the following scheme



Noradrenaline exerts little effect upon the concentration of lactic acid in the peripheral blood (Bearn et al)

(g) *General Metabolism* Oxygen consumption is increased by from 20 to 40 per cent, and CO_2 production by from 30 to 50 per cent, the respiratory quotient is therefore raised In man the increase in the basal metabolic rate occurs within a short time after the subcutaneous injection of 0.5 cc. of a 1:1000 solution, the temperature of the muscles rises The effect of adrenaline upon heat production is known as its *calorigenic* action, it does not occur after removal of the liver (Soskin)

Boothby and Sandiford attributed the latter effect largely to stimulation of cellular oxidations throughout the tissues of the body generally, and in a minor degree to the hyperglycemia ("carbohydrate plethora"), and the increased utilization of carbohydrate, that is, to the specific dynamic action of glucose The extent to which each of these factors contribute to the increased heat production is a controversial question, but the consensus favors increased carbohydrate utilization as the predominant factor (see Griffiths) Cutaneous vasoconstriction leading to diminished heat loss and its effect upon metabolism, as well as increased muscular tone, are probably contributing factors The calorigenic effect is not brought about through an action upon the thyroid since it is obtained after thyroidectomy Moreover the rise in metabolic rate commences within a few minutes and returns to normal within 2 hours or so, whereas the effect of thyroxine upon heat production does not commence for some hours and is prolonged for several days (p. 823)

(h) *Other effects of adrenaline (U.S.P.)* (1) Secretion of saliva, (2) lacrymation, (3) sweating in such animals as horses and sheep, but in most other animals the sweat glands, though innervated by the sympathetic, are not excited by adrenaline (see p. 734) Sweating can be evoked in most human subjects by adrenaline or noradrenaline, (4) contraction of the spleen (stimulation of the smooth

muscle of its capsule and trabeculae) and consequent increase in the blood volume and in the red cell count (p. 70), (5) increase in the coagulability of the blood, (6) small doses increase the flow of urine as a result of constriction of efferent glomerular vessels, larger doses constrict both afferent and efferent vessels and through reduction of the renal blood flow diminish the urinary flow (p. 452), (7) a fall, sometimes preceded by a rise, in the potassium of the blood, (8) contraction of melanophores of certain cold-blooded animals, e.g., frog and horned toad (Redfield),² (9) increased output of the adrenocorticotrophic, thyrotrophic, and "ovulation" hormones of the pituitary gland (p. 886), (10) affects the transmission in nerve, (11) enhances or depresses, according to the

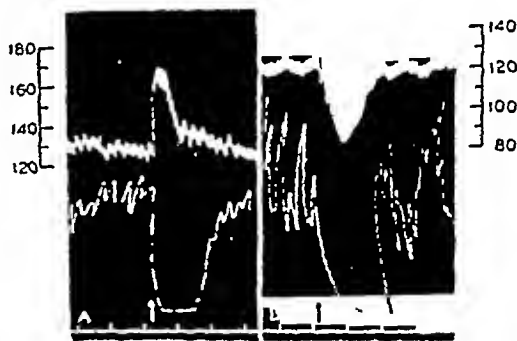


FIG. 594 Showing the effects of adrenaline upon the blood pressure and the uterus (*in situ*) of a pregnant cat before and after the administration of dibenamine. The intravenous injection of 2.5 mg. per kg. is indicated by the arrows (After Nickerson and Goodman)

dosage, the action of acetylcholine at synaptic junctions (p. 1102)

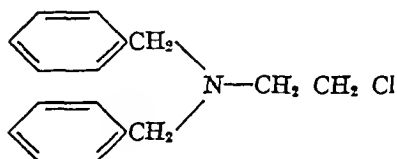
THE EFFECTS OF CERTAIN SUBSTANCES UPON THE ACTION OF ADRENALINE Ergotoxine³ or ergotamine (derivatives of ergot) annuls the excitatory (motor and secretory) responses of adrenaline or of sympathetic stimulation, but inhibitory effects, e.g., vasodilator, are

² The contraction of the melanophores of the horned toad which results from adrenal administration occurs when the animal becomes excited. Since in the latter instance the effect is abolished if the circulation to the melanophores is occluded but not when the skin nerves are cut, the effect must be due to a chemical substance—presumably adrenaline—in the blood.

³ Ergotoxine is made up of three alkaloids, ergocornine, ergocristine, and ergocryptine, derivatives of lysergic acid. These compounds, when hydrogenated to dihydroergocornine, -cristine and -cryptine, are much less toxic than ergotoxine, and permit their experimental use in man. Of the three, the action of dihydroergocornine has been the most fully investigated. It blocks the action of both adrenaline and noradrenaline.

not interfered with.⁴ For this reason certain effects of adrenaline administration or of sympathetic stimulation appear to be reversed by the ergot alkaloids. Thus, as already mentioned, only a depressor effect is obtainable with adrenaline if ergotoxine has been administered previously. Also, after ergotoxine, adrenaline causes expansion of the melanophores of the frog instead of the usual contraction. The hyperglycemic response to adrenaline is abolished after ergotoxine. *Apocodein* is another drug which reverses or annuls some of the effects of adrenaline whereas *cocaine* enhances its vasoconstrictor, cardiac and pupillary reactions. In their effects upon the capillaries, blood pressure and bronchioles, *histamine* and adrenaline are antagonistic, and adrenalectomized animals show an increased susceptibility to histamine administration. The blood-concentrating effect of histamine administration is lessened or prevented by a previous injection of adrenaline. There is also some evidence that histamine increases the output of adrenaline from the medulla.

Other antagonists of adrenaline and of sympathetic effects have been discovered more recently. *Dibenamine* (N,N-dibenzyl-β-chloroethylamine) is especially effective in blocking adrenaline or sympathetic action (see fig 594). This drug though it reverses the pressor action of adrenaline differs from ergotoxine in that it does not prevent the hyperglycemic action of adrenaline. Nor are the inhibitory effects of adrenaline upon the virgin cat's uterus and the intestine altered by dibenamine. The formula of dibenamine is shown below.

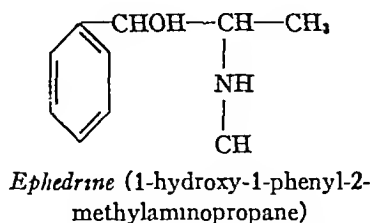


Fourneau compounds, such as the *benzodioxanes*, 933F and 883F, also block adrenergic nerves, as well as the actions of adrenaline and noradrenaline. Besides these peripherally acting (adrenergic blocking) agents others which abolish ganglionic transmission, such as nicotine, tetraethyl ammonium and the methonium compounds, may be mentioned here. Nicotine acts by depolarizing the ganglion cell, which it first stimulates and then paralyzes. The methonium compounds (penta-, hexa-, and decamethonium) bring about their blocking action in a different way, namely, by competing with naturally formed acetylcholine for "possession" of the ganglion cell, that is, for the specific chemical receptors of the cell. Their ability to compete with acetylcholine depends upon their resembling the latter chemically, being also quaternary nitrogen compounds.

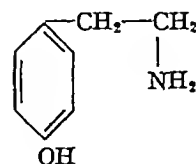
⁴ The abolition of the effect of sympathetic stimulation, is called a *sympatholytic* action, while a drug which abolishes the excitatory action of adrenaline is called *adrenolytic*. The term *adrenergic blocking agent* may be applied to a drug which has either an adrenolytic or a sympatholytic action, or both.

The methonium compounds have come into clinical use for the reduction of the blood pressure (ch 16), and for their curari-like action.

ADRENALINE-LIKE SUBSTANCES *Ephedrine* was isolated by Chen in 1924 from the Chinese plant *ma huang*. Its chemical resemblance to adrenaline may be seen from the formula.

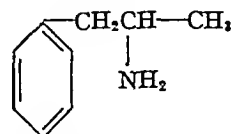


This alkaloid is closely similar to adrenaline in its action, causing bronchiolar relaxation, vasoconstriction, hyperglycemia, inhibition of intestinal muscle and excitation of other smooth muscle. It has also a calorigenic action, its pressor effect (Curtis) is reversed by ergotoxine and reduced by cocaine. Ephedrine is about 1000 times less powerful than adrenaline, but its effects are more prolonged. Because of its more lasting effect it is often combined with adrenaline in commercial preparations. Unlike adrenaline it is active when taken by mouth, 50 to 150 mg causing a pronounced rise in blood pressure. According to the theory of Gaddum and Kwiatkowski, ephedrine exerts its action after a fashion analogous with that by which eserine enhances the action of acetylcholine (p 1098), namely, by inhibiting the oxidative deamination of adrenaline by amino-oxidase. *Tyramine* is a sympathomimetic amine formed by the action of bacteria upon protein (decarboxylation of tyrosine). It is also secreted by the salivary glands of certain molluscs. It is chemically allied to adrenaline and resembles the latter in several of its actions, e.g., it elevates the blood pressure and stimulates the uterine muscle. Its structural formula is



Tyramine (1-(4'-hydroxyphenyl)-2-aminoethane)

Amphetamine or *benzedrine* is a compound closely allied chemically to adrenaline, as shown in the following formula.



Amphetamine (1-phenyl-2-aminopropane)

This sympathomimetic amine has also a stimulating action upon the mental processes somewhat like that of caffeine, inducing wakefulness, postponing mental fatigue and relieving nervous depression.

A number of other sympathomimetic amines have been discovered, e.g. phenylethylamine, neosynephrine, 2 naphthyl (1st)-methyl madazoline (privine), 2-aminoheptane (tuamine). For the chemistry and action of such compounds the reader is referred to a paper by Beyer.

The skins of certain toads secrete substances (*bisfotensine* and *bisfotenedine*) similar in action to adrenaline, the secretion of these substances is increased during nervous excitement.

THE EMPLOYMENT OF ADRENALINE IN MEDICINE. Adrenaline is used (a) To staunch bleeding from accessible mucous surfaces (epistaxis, hematemesis) or from the skin. On account of its pressor effect it is of no value and may do harm in pulmonary hemorrhage and other forms of internal bleeding. (b) To shrink mucous surfaces, especially of the nose, and is then frequently used in combination with ephedrine. (c) To relax the muscle of the bronchioles in asthma, the relief following its use is very striking. (d) To combat certain allergic conditions, e.g., serum sickness, and to antagonize the effects of the histamine like substances supposedly liberated in the skin in such conditions as urticaria and giant edema. (e) To stimulate the respirations or to resuscitate the heart which has ceased to beat, in the latter event it may be given by injection into the cardiac musculature. Adrenaline is of value in syncope due to Stokes-Adams disease (p 229). It should be mentioned that the administration of adrenaline during chloroform anesthesia is attended by the danger of inducing ventricular fibrillation (p 237). (f) To enhance and prolong the action of cocaine and similar local anesthetics, the vasoconstriction thus induced retards the absorption of these substances, thus prolonging their effects and reducing their toxicity. It is also used as an aid in spinal anesthesia. (g) Noradrenaline has been used in certain hypotensive states for its vasoconstrictor action.

BIOLOGICAL TESTS FOR ADRENALINE

(a) *Intestinal segment* The fluid suspected of containing adrenaline (blood, serum, etc.) is applied to a segment of surviving rabbit intestine. Inhibition of the intestinal movements will result if the fluid contains adrenaline. An estimation of the adrenaline concentration in the fluid may be arrived at by matching its action with that of an adrenaline solution of known dilution.

(b) *A segment of the non pregnant uterus of the rabbit* may be employed in a similar manner. This test object is stimulated in the presence of adrenaline.

(c) *Isolated arterial rings* A ring of a small artery known to react to adrenaline by constriction is sus-

pended in Locke's solution. The arterial ring is attached by a thread to a lever applied to a smoked surface. The fluid to be tested is added to the bath in which the arterial ring is immersed.

(d) *The denervated iris* Removal of the superior cervical ganglion renders the pupil highly sensitive to the dilator action of adrenaline (the paradoxical pupillary reaction of Meltzer). The test fluid is instilled into the conjunctival sac of an animal so prepared.

(e) *The caval pocket method of Stewart and Rogoff* This is an auto-assay method. The liberation of adrenaline is demonstrated through its action upon the denervated iris. The inferior vena cava is clamped below the entrance of the adrenal veins and again just below the diaphragm. All veins entering this caval pocket, except the adrenal veins, are ligated. After a measured period of time the upper clamp is removed, dilatation of the pupil results if adrenaline has been secreted in the interval between the closure and opening of the upper clamp. In quantitative determinations the lower clamp is removed and the blood collected through a cannula inserted into the vein. The concentration of adrenaline in the collected blood is then estimated from tests upon a segment of beating intestine.

The denervated heart (described on p 835)

THE SECRETORY ACTIVITY OF THE ADRENAL MEDULLA

The secretion of the chromaffin cells of the adrenal, which enters the general circulation via the adrenal vein, contains both adrenaline and noradrenaline, but the proportions of the two in the circulation are unknown. The discharge of the secretion from the gland appears to be entirely under nervous control, denervation of the adrenals results in suppression of their secretion. That the secretion of adrenaline is dependent upon nerve impulses has long been recognized, but the liberation of noradrenaline has been demonstrated more recently by Bulbring and Burn, who detected it in the circulation of the eviscerated cat with intact adrenals after stimulation of the splanchnic nerve.

It is not believed that significant amounts of the secretion of the adrenal medulla enter the circulation under ordinary resting conditions. Stewart and Rogoff some years ago determined in animals, by the caval pocket method, the rate of adrenaline secretion and estimated that during bodily rest the amount of the hormone in the circulation was not greater than from 1 2,000,000,000 to 1 1,000,000,000 (Rogoff). Such a high dilution has no demonstrable effect in the intact normal animal. On the other hand, Cannon and Rapport found that in states of stress (see emergency func-

tion, below) or during splanchnic stimulation, the output of adrenaline amounted to from 0.003 to 0.004 mg per kilogram per minute. Reduction or exhaustion of the adrenaline content of the gland results from stimulation of the great splanchnic nerve, puncture of the floor of the fourth ventricle or reflexly from excitation of a sensory nerve. The adrenaline liberation caused by one or other of these procedures may cause an increase in metabolism of over 20 per cent. Anesthesia, asphyxia, strenuous exercise, a fall in blood pressure or any state of stress. The administration of such drugs as strychnine, nicotine and morphine also cause the secretion of adrenaline if the nerves to the gland are intact, but not after the gland has been completely denervated. Massage of the adrenal also causes the passage of adrenaline into the blood stream. Cannon and Rapport place the center for the control of adrenal secretion in the upper part of the floor of the fourth ventricle. This is probably not the highest center since stimulation of the hypothalamus (paraventricular nuclei) will cause a discharge of adrenaline.

THE RÔLE OF THE ADRENAL MEDULLA IN THE BODY

There seems little doubt that the effects following the injection of adrenaline indicate the physiological actions of the secreted hormone in the living body. That adrenaline is a true hormone, and not merely a pharmacological agent, is supported by the following facts:

(a) The almost perfect correspondence between the effects of adrenaline administration and those resulting from the stimulation of certain sympathetic nerves, (b) the common origin of the adrenal medulla and the sympathetic nervous system, (c) discharge of adrenaline into the blood stream under the experimental conditions mentioned above, and (d) the results of researches to be immediately described.

The adrenal medulla is not essential to life. In animals one adrenal may be removed completely and the medulla of the other excised without any apparent ill effect—the animal survives the operation indefinitely.

The emergency theory of adrenal function

Cannon and his colleagues have furnished convincing evidence that the medulla liberates its secretion in significant amounts only under conditions which call for unusual effort on the part of the body to perform work, to prevent changes in

its internal environment or to resist threatened dangers. In such times of stress the medullary secretion, it is believed, reinforces the sympathetic nervous system. Through this hormonal-nervous cooperation the several bodily reactions associated with such states of emergency are raised to maximal efficiency. Cannon and his associates employed the denervated heart as an indicator of adrenaline liberation. The operation for denervation comprises section of the vagi and removal of the stellate and second thoracic ganglia of the sympathetic chain; the heart is thus completely isolated from nervous control. Since in their experiments any effect due to a change in the temperature of the blood was excluded, a pronounced acceleration of a heart so prepared was taken to be the result of a chemical substance carried in the blood stream. Fright, rage, pain, asphyxia, anesthesia, muscular activity, exposure to cold, stimulation of a sensory nerve and several other conditions, caused within 10 seconds an increase in heart rate of from 20 to 40 beats per minute. Removal of the adrenals, their denervation, or ligation of the adrenal veins, prevented this effect. The conclusion, therefore, is justified that the various conditions mentioned cause the reflex liberation of the medullary hormone. The denervated heart responds to as little as 1 part of adrenaline in 1400 million parts of blood. In some of Cannon's experiments, cats were frightened by the barking of a dog, the rate of the denervated heart increased by from 15 to 30 beats per minute. The cardiac acceleration was accompanied by pupillary dilation, erection of the hairs and spitting. When motor activity, e.g., struggling in the animal holder, accompanied the emotional excitement, the cardiac acceleration was more pronounced (40 to 80 beats per minute). Even minor muscular movements without emotion, e.g., extending the legs, walking or turning the body caused an acceleration of from 5 to 20 beats.

The hyperglycemia and glycosuria resulting from emotional excitement in man and animals is probably associated with the discharge of adrenaline, since it has been shown that the continued rise in blood sugar which occurs during the emotional reactions (sham rage) following removal of the cerebral cortex (p. 1027) is dependent upon the adrenals. The blood sugar continues to rise after this operation, though the glycogen stores of the liver are removed from nervous control by sectioning the hepatic nerves. On the other hand, the effect does not occur after removal of the adrenals,

even though the hepatic nerves are intact. In those instances in which the emotional state does not follow the operation of decortication, the hyperglycemic effect also fails to appear. Puncture of the floor of the fourth ventricle or stimulation of the adrenal nerves also causes hyperglycemia after the hepatic nerves have been previously cut.

Though direct evidence is difficult to obtain it is reasonable to assume that besides the effects mentioned above, adrenaline when secreted into the blood stream brings about other actions which we have seen to be characteristic of its action when injected.

A recapitulation of the actions of the sympatho-adrenal system will show how important these several actions are in fitting an animal for defense or flight, for attack or pursuit. (1) The rise in general blood pressure accompanied by dilatation of the vessels of the contracting skeletal muscles, and of the coronary arteries, and the increased force and output of the heart, raise the circulatory system to a state of maximal efficiency. (2) Hyperglycemia indicates the mobilization of the carbohydrate stores of the liver, thus an adequate supply of fuel for the active muscles is ensured, muscular fatigue occurs less readily. (3) Increased oxygen capacity of the blood is brought about by the discharge of red cells from the spleen. (4) Bronchiolar dilatation and an increase in the rate and depth of respiration permit an increased oxygen intake to supply the tissue cells, at the same time the level of oxygen consumption of the latter is raised. (5) Shortened coagulation time of the blood lessens the danger from hemorrhage. (6) Finally, the emotional manifestations of man and the fighting attitudes or defense reactions of various animals are sympatho-adrenal effects, e.g., pupillary dilatation, and, possibly, the startled expression due to contraction of Mueller's orbital muscle which retracts the upper eyelid, cutaneous vasoconstriction, acceleration of the heart, contraction of smooth muscle in the skin causing "gooseflesh" in man, and the erection of the hairs, quills or feathers of animals, sweating, salivary secretion (cat),⁵ and the color changes of some cold-blooded animals.

⁵ It is an interesting and perhaps a significant fact that in the cat, in which spitting is a defense reaction, sympathetic stimulation causes a profuse watery flow of juice from the salivary glands. In other animals sympathetic stimulation causes a scanty flow of viscid saliva, a watery secretion being caused by parasympathetic excitation.

The tonus theory of adrenal function

It has been suggested that the medullo-adrenal secretion maintains the sympathetic nerve endings in a state of sensitivity or tone and that the height of the normal blood pressure is dependent upon the continuous discharge of the hormone into the blood. Low blood pressure has been ascribed to adrenaline deficiency (so-called hypoadrenalemia) and essential hypertension to the liberation of adrenaline in excess (hyperadrenalemia). But the evidence against such conjectures is conclusive, the theory has been entirely discredited. For example, if one adrenal is excised and the other curetted or burnt away, no fall in blood pressure occurs, so long as a sufficient amount of cortical tissue is left intact. Furthermore, we have seen that adrenaline does not increase the peripheral resistance (which is the basis of essential hypertension), and neither noradrenaline nor adrenaline is found in greater than usual amounts in the adrenals of hypertensive subjects. There are, however, certain types of hypertension, namely those associated with pheochromocytoma or adrenal hyperplasia which are due to hypersecretion of the hormones of the adrenal medulla.

Pheochromocytoma and hyperplasia of the adrenal medulla

Pheochromocytoma is a tumor composed of chromaffin tissue arising either in the adrenal medulla itself or in an outlying collection of chromaffin cells (paraganglia, Zuckerkandl's organ). The secretion of one of these growths, or of a simple hyperplasia of the adrenal medulla, is a rare cause of hypertension, though not as rare as it once was thought to be. The hypertension may be due to the presence in the circulation of unusual amounts of adrenaline, or of noradrenaline, or of both hormones, of which the tumor or hyperplastic medullary tissue contains excessive amounts. Noradrenaline rather than adrenaline is usually in greatest concentration and may constitute 90 per cent of the gland's hormone content. Both hormones have been found in excess in the urine of subjects of this type of arterial hypertension. The hypertension may be paroxysmal, with sometimes a violent onset, or the excessive secretion of the hormones may be continuous, the high blood pressure then being sustained and indistinguishable, clinically, from essential hypertension. Hyperglycemia and glycosuria, especially when the disease is paroxysmal, may occur, pallor, rapid pulse and profuse sweating usually accompany the attack. Peripheral circulatory collapse may ensue during a paroxysm which is thought to be analogous to the shock in animals which follows the injection of large doses of adrenaline, or hypotension and circulatory collapse may result from surgical removal of the tumor. This is attributed to the depressed tone of the vasomotor center, made evident only after the abolishment of the excessive adrenal secretion. Adrenal hypertension can be distinguished from essential and other hypertensive states by means of Gold-

enberg's benzodioxane test. Benzodioxanes temporarily abolish, through their adrenergic blocking action, hypertension of adrenal origin, but not that due to other causes

SYMPATHIN

It has been mentioned that the *prompt* acceleration of the denervated heart does not occur in adrenalectomized animals during excitement, sensory nerve stimulation, etc. Cannon and his colleagues found, however, that a *slowly developed* acceleration of the denervated heart occurred during excitement or muscular activity though the adrenals had been extirpated. The increase in heart rate took about a minute to develop, reached its maximum in about 3 minutes and then gradually subsided. Its occurrence was not prevented by the removal of all accessory adrenal tissue, by hypophysectomy or by the excision of the thyroid, parathyroids or gonads, or by denervation of the liver. It was abolished, however, by removal of the sympathetic chains. It was also found that the characteristic slow acceleration of the heart occurred when the lower abdominal sympathetic chain was stimulated. The latter nerve contains fibers supplying the smooth muscle of the skin which are responsible for the erection of the tail hairs. Secretion of the denervated salivary glands, contraction of the nictitating membrane, a rise in blood pressure and of blood sugar also resulted from the stimulation of the abdominal sympathetic. These effects as well as the cardiac acceleration occurred though the cord had been divided in the thoracic region, and the sympathetic chain above this level removed. A material originating in the hind part of the animal was evidently conveyed in the blood stream to the heart and other structures mentioned. Blocking the blood flow returning from the area supplied by the stimulated nerve or removal of the patch of skin prevented the cardiac response. As a result of these researches Cannon and his associates concluded that during sympathetic stimulation a chemical substance resembling adrenaline in its action was liberated from the sympathetic endings supplying the smooth muscle of the skin. They named this substance *sympathin*.

Cannon and Rosenblueth found that sympathin and adrenaline showed certain differences in action. Preparations of adrenaline (e.g., U.S.P. epinephrine) exhibited both excitatory and inhibitory actions, whereas, there appeared to be two types of sympathin, one, which they named *sympathin E* (excitatory), was formed at the terminals of some

sympathetic nerves, such as vasoconstrictors, while the other type, called *sympathin I* (inhibitory), was produced at other nerve terminals, such as those in the small intestine. But sympathin E was found to be not purely excitatory, and sympathin I not purely inhibitory. This rather confusing state of affairs has been clarified by the discovery of the existence of two hormones of the adrenal medulla. Sympathin E is now considered to be noradrenaline, as first suggested by Bacq; sympathin I is thought to be adrenaline itself. von Euler proposes that the qualifying letters E and I should be substituted by N (for noradrenaline) and A (for adrenaline). It is likely that noradrenaline rather than adrenaline is the predominant humor liberated at nerve endings.

THE ADRENAL CORTEX

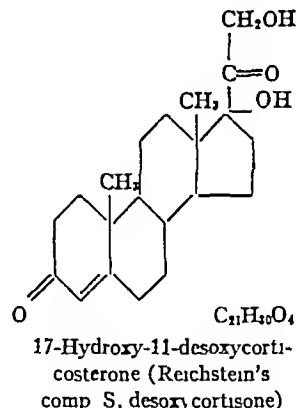
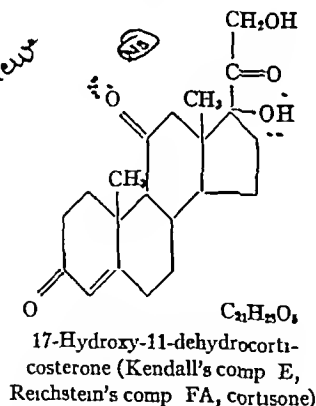
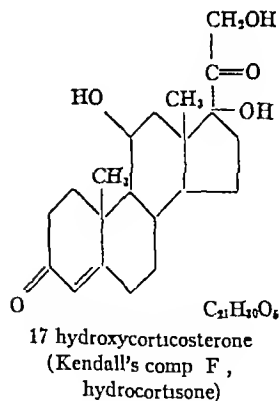
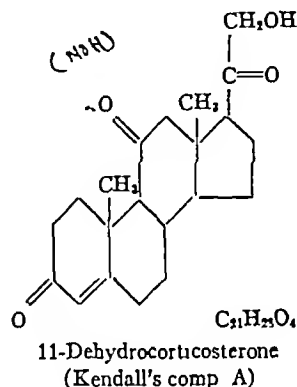
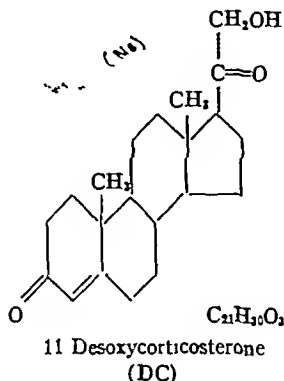
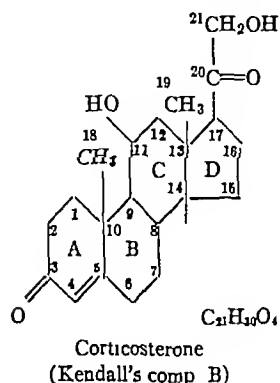
The cortex of the adrenal, unlike the medulla, is essential to life. Removal of more than about five-sixths of this part of the adrenal causes death within a few days. Removal of the interrenal body of Elasmobranch fishes (p. 827) is also fatal. Stewart and Rogoff found that the average survival time of dogs after complete double adrenalectomy was 10 days and the maximal time 15 days. They observed that pregnant animals or those in heat (pseudopregnancy) actually survived much longer. For this reason, they suggested that some substance formed at these times substituted for the cortical hormone. This observation has been amply confirmed, and it has been shown further that the survival of adrenalectomized dogs and rats is extended by the administration of progesterone (p. 881). Stewart and Rogoff showed also that the lives of completely adrenalectomized animals could be prolonged by the injection of a cortical extract combined with the transfusion of saline. Hartman too obtained an extract which he called *cortin*. This was capable of definitely prolonging the survival time of adrenalectomized animals. Swingle and Pfaffner in 1930 extracted by means of lipid solvents a very potent substance from the cortex which counteracted the effects of adrenal deprivation. When treated with this preparation adrenalectomized animals survived indefinitely, it also proved highly successful in the treatment of Addison's disease.⁶

⁶ Dogs are employed for assaying the potency of the extract. A dog unit (D.U.) is defined as the minimal daily quantity per kg. of body weight which will maintain for from 7-10 days an adrenalectomized dog in normal condition, as judged by the blood, non-protein nitrogen level and body weight.

AN OUTLINE OF THE CHEMISTRY OF ADRENOCORTICAL PRINCIPLES

These are now often referred to as *adrenal corticoids*. Those which act mainly upon the metabolism of water and salt have been named *mineralocorticoids*. Those which affect the metabolism of carbohydrate are called *glucocorticoids*. Between 20 and 30 steroid compounds have been isolated from the adrenal cortex, but most of these are physiologically inactive. The formulae of the six best known of the crystalline physiologically active compounds

It is absent, or obtainable only in insignificant amounts, from adrenal extracts. This synthetic compound which is principally concerned in the metabolism of water and salt (Na and K) is administered in the form of the acetate, either intramuscularly or as pellets implanted subcutaneously. A material called the *amorphous fraction* remains in the adrenal extract after the foregoing crystalline compounds have been removed. It contains 90 per cent of the physiological potency of the whole cortical extract.



—corticosterone, desoxycorticosterone, 11-dehydrocorticosterone, 17-hydroxycorticosterone, 17-hydroxy-11-dehydrocorticosterone and 17-hydroxy-11-desoxycorticosterone—are given above.

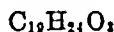
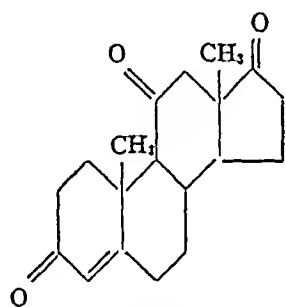
Corticosterone was isolated by Reichstein and his associates in 1937, and by Kendall et al in the same year, and was later synthesized by the former group of workers. It is also known as Kendall's compound B. Desoxycorticosterone, which lacks the hydroxyl group at position 11, was later synthesized by Steiger and Reichstein from stigmasterol.

Recently a steroid with up to 100 times the mineralocorticoid activity of desoxycorticosterone has been isolated from the amorphous fraction. It is known as *electrocortin* or, since it has an aldehyde group at C18, as *aldosterone*. Otherwise it is identical chemically with corticosterone. A substance with a similar action has been found in adrenal blood of the dog and the monkey.

Besides compounds exhibiting the specific actions of the adrenal cortex, others with estrogenic,

progestational, or androgenic properties have been isolated from adrenal extracts, these are *cstrone*, *progesterone*, *17-“β”-hydroxyprogesterone*, *Δ⁴-androstenedione-3,17*, and *adrenosterone* (*Δ⁴-androstenedione-3,11,17*) The last four of these compounds are androgenic, and the last three have a ketone group in position 17, and are known as adrenal neutral *17-ketosteroids* (17-KS) The 17-ketosteroids excreted in normal urine are, *androsterone*, *dehydroisoandrosterone* and *11-hydroxyisoandrosterone* Cortical steroids of both the 11-oxy, and the 11-desoxy series are also excreted in the urine These and the 17-KS are, as mentioned below, excreted in increased amounts in adrenal tumor or after the administration of the adrenocorticotrophic hormone

The most active androgen in the adrenal cortex is adrenosterone Its formula is given below



Adrenosterone

The daily urinary excretion of neutral 17-ketosteroids averages 15 mg (65 international units) and 10 mg for men and women, respectively, and 1 mg for children under 7 years of age In women all is derived from the adrenal cortex, there being no evidence of any being of ovarian origin In men two-thirds of the urinary 17-KS are of adrenal origin, the remainder being furnished by the testes The adrenal 17-ketosteroids are possibly metabolic derivatives of those corticoids having an OH group at position 17, such as, compounds E and F Presumptive but inconclusive evidence for the origin of the adrenal 17-KS from adrenal steroids with an hydroxyl group at position 17 is that their excretion is reduced in Addison's disease and in pituitary cachexia, but is increased in cortical tumor or hyperplasia, and in states of stress The conversion of 17-hydroxylated steroids to 17-keto steroids is readily accomplished *in vitro*, and the administration of adrenocorticotrophin in man causes a considerable increase in 17-ketosteroid excretion Alternately it has been suggested that the adrenal 17-ketosteroids are produced from cholesterol as are also, most probably, the corticoids, but through separate and independent pathways

(Morris) Estrone is an acid 17-ketosteroid, and is not detected by the method of estimation employed, it is, therefor, not included as part of the total 17-ketosteroid excretion (See also p 905)

The precursor of corticoids, cholesterol and ascorbic acid depletion by ACTH

There is good evidence that cholesterol is the parent substance of the cortical hormones The cholesterol content of the adrenal cortex is reduced by 50 per cent by the injection of the adrenocorticotrophic hormone, by trauma, or during states of stress (Long and associates) The greatest reduction in the cholesterol content was reached in 5 or 6 hours after the injection of ACTH, and required some 24 hours to be restored to normal This is a specific effect of the hormone upon the adrenal cortex, for no similar reduction of the cholesterol of other organs was observed Increase of liver glycogen accompanies the diminution in adrenal cholesterol, which is taken as an indication of the release of glucocorticoids Evidence secured through the use of isotopically labelled carbon (C^{13}) also points clearly to cholesterol as the material from which the cortical steroids (as well as the bile salts) are elaborated Chaikoff has demonstrated the synthesis of cholesterol by slices of beef adrenal incubated with labelled sodium acetate

The ascorbic acid content of the adrenal is also diminished by 50 per cent or more by ACTH or by trauma In the rat, which can synthesize it, the vitamin after its reduction commences to reaccumulate within 3 hours, and its concentration in the gland is restored to normal or above within 12 hours In the guinea pig, on the contrary, which depends upon the diet for its supply of the vitamin, no accumulation occurs by this time The significance of these results with respect to adrenal cortical function is unknown, for corticoids are elaborated in the absence of the vitamin, that is, in scorbutic animals

The natural cortical hormone

Though some uncertainty still remains, the evidence points to compound E or F, or both, as being normally secreted by the human adrenal cortex In the first place the administration of adrenocorticotrophic hormone (ACTH), the natural stimulant of adrenocortical activity, duplicates the effects induced by the administration of compound E or F itself The effect of ACTH is abolished, of

course, after adrenalectomy. The actions of E or F or of ACTH approach more closely to what might be expected of a natural hormone than does desoxycorticosterone, which fails to correct fully the defects of adrenal insufficiency, namely, the tendency to hypoglycemia during fasting, and the abnormalities of the electroencephalogram. It also exerts an abnormally great antidiuretic response and does not induce eosinopenia, a characteristic action of cortisone and ACTH. These principles however exert little effect upon salt and water metabolism.

But more direct evidence can be cited to support the belief that cortisone (E) and hydrocortisone (F) are the natural hormones. Only cortisone has been isolated from normal human urine. Hydrocortisone has been found in large amounts in the urine of patients with adrenocortical tumors. Compounds A and B (dehydrocorticosterone and corticosterone) are probably the immediate precursors, respectively, of compounds E and F (Kendall).

There appears, however, to be a species difference in the proportions of the active steroid hormone content in the adrenals, as well as of the compounds secreted. Hechter, in perfusion experiments of surviving beef adrenals, found that the addition of ACTH to the perfusion fluid caused an increase in the effluent of compounds B (corticosterone) and F, but no increase of E.

The manifestations of cortical deficiency in animals

An animal which has been completely adrenalectomized shows the following features during the short period of its survival. Loss of appetite (particularly for fats), vomiting, diarrhea, rapid loss of weight, muscular weakness and prostration, pronounced diuresis until the later stages, fall in body temperature, hypotension, and a reduction in the basal metabolic rate of about 20 per cent (fig 59.5). The metabolic response to a cold environment is less than that of a normal animal, and any form of stress is poorly met. The blood becomes concentrated (loss of plasma water) and shows a fall in sodium and sugar, and a rise in non protein nitrogen, phosphate, calcium and especially of potassium. The urinary excretion of sodium is increased (p 842) and that of potassium reduced (see fig 59.6). There is reduced excretion of urinary nitrogen and other signs of renal failure. Though there is spontaneous diuresis the elimination of ingested water is defective. The glycogen

store of the liver and muscles are diminished. These changes, as well as the effect upon the electrolyte concentrations of the blood, are detectable within 42 hours following adrenalectomy. Post-mortem examination frequently shows congestion of the gastrointestinal tract and pancreas.⁷

The foregoing picture can be entirely prevented by the administration of an extract of the cortex and even comatose animals can be restored to health, heat, pregnancy and lactation then proceed normally in adrenalectomized bitches, and growth is maintained in adrenalectomized puppies.

The cortex quite evidently is concerned with water metabolism and with the metabolism of potassium and sodium (see also ch 35). In adre-

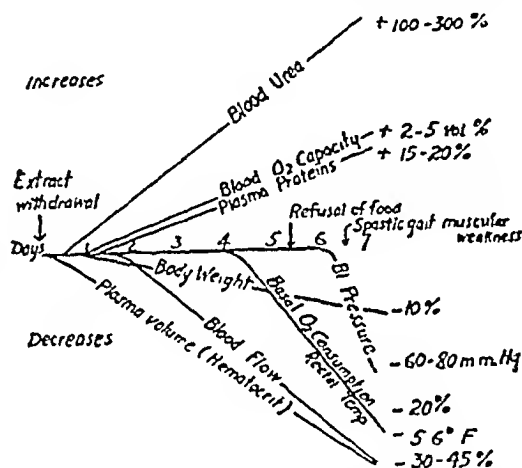


FIG 59.5 Characteristic changes in metabolism, circulation and blood concentration in a group of adrenalectomized animals following withdrawal of extract. (After Harrop and associates)

nalectomized animals as well as in the adrenal insufficiency of man, a marked reduction in total base occurs which is due mainly to the loss of sodium. The potassium concentration of the serum is raised above the normal level, the rise being due to reduced excretion by the kidney as well as to leakage from the tissue cells into the extracellular fluids. The loss of sodium due apparently to diminished reabsorption by the renal tubules is accompanied by an increased elimination of water. Marked dehydration results. The administration of sodium chloride to adrenalectomized animals

⁷ Pigmentation as seen in man has not been observed in lower animals, though, according to Hartman in white cats a grayish discoloration of the skin may be noted. Bronze freckles have also been described in adrenalectomized monkeys and pigmented spots in the buccal mucosa of dogs.

and a reduction in the potassium intake exert a remarkably beneficial effect. On the other hand, the withdrawal of salt from the diet increases the severity of the condition. The sodium rather than the chloride ion is the important factor.

The reduction of blood sodium was observed originally by Bauman and Kurland, and the beneficial effects of transfusions of saline in sustaining adrenalectomized animals were described by Stewart and Rogoff and by Hartman, but the importance of sodium loss in the development of the symptoms was first pointed out by Loeb and his associates and by Harrop. The value of a low potassium content of the diet in the survival of adrenalectomized animals was shown by Allers and his associates. Completely adrenalectomized animals can be maintained in good health without treatment by a cortical hormone if placed upon a diet low in potassium but high in sodium chloride and, in order to prevent the acidosis which otherwise develops, containing sodium citrate. The tendency to hypoglycemia is controlled by a high carbohydrate diet. These measures, however, do not completely restore an animal to a physiological state, for they are unable to withstand stress or to perform work with full efficiency.

It is claimed that increased susceptibility to certain toxic agents, e.g., histamine, morphine, diphtheria toxin, etc., is another manifestation of adrenal insufficiency. The adrenal cortex also appears to be concerned in some way with the maintenance of the nutrition of certain nervous tissues; degenerative changes have been noticed in autonomic ganglia (cervical, stellate and celiac) following adrenalectomy, and almost complete loss of function of the vasomotor and cardiac accelerator nerves in adrenalectomized animals was demonstrated years ago by Elliott and more recently by Cleghorn and associates. The intestinal absorption of water and electrolytes and the diuretic response to water drinking are greatly reduced in adrenalectomized animals. The relation of the adrenal cortical hormone to the development of renal hypertension is described in chapter 16, and its rôle of the cortex in lactation is discussed in chapter 61.

The physiological actions of the cortical extract and of its different fractions

All the active cortical steroids whose formulae are shown on p. 838 have a double bond between positions 4 and 5; it appears to be the *sine qua non* of physiological activity, a compound being rendered inert by hydrogenation in this position. An oxygen at position 21 is also necessary for any

marked degree of cortical activity. Progesterone, which lacks it, has only a very weak cortical action. An oxygen atom at position 11 either as $=O$ or $-OH$ as in compound E or F endows a compound with the power to act upon carbohydrate, protein and fat metabolism; a hydroxyl group in addition at position 17 enhances this effect. The $-OH$ group in this latter position is also essential for the antirheumatic and other allied actions of compounds E and F. The compounds A and B which are identical in chemical structure with the latter, except that the $-OH$ at position 17 is lacking, have no antirheumatic action. Desoxycorticosterone in which no ketone or hydroxyl group is present at position 11, and the hydroxyl group is absent as well from position 17, also lacks antirheumatic effect.

The fundamental action of cortical steroids (glucocorticoids) upon carbohydrate and protein metabolism is the catabolism of tissue protein, the carbon part of the amino acids thus released being utilized for the new formation of glucose (gluconeogenesis) and the accompanying storage of glycogen in the liver. Due to the stimulation of gluconeogenesis the glucocorticoids exert an anti-insulin action. If inadequate amounts of these hormones are secreted (as in adrenal insufficiency) hypoglycemia during fasting is likely to result. The adrenalectomized animal is depancreatized as well, therefore does not exhibit the usual hyperglycemic effect of the latter operation.

The glucocorticoids cause the mobilization of fat from the fat depots and increase the fat content of the liver. Nothing more need be said here concerning the role played by the adrenal cortex in the metabolism of carbohydrate and fat, since these subjects have been dealt with in chapters 49 and 50.

Corticosterone and other compounds with an oxygen atom at position 11 reduce muscular fatigue, whereas desoxycorticosterone is inert in this respect; the glucocorticoids on the other hand exert only a slight effect on salt and water metabolism.

The restoration of normal kidney function is exhibited by desoxycorticosterone but to the greatest degree by the amorphous fraction. Either of these compounds causes a prompt reduction in the non-protein nitrogen of the blood. The most outstanding action of desoxycorticosterone is upon salt and water metabolism; it increases the plasma volume and the concentration of sodium in the body fluids,

but reduces that of potassium. It reduces the intracellular concentration of potassium and increases that of sodium, and with continuing dosage the concentration of Na may nearly equal that of K. The administration of potassium almost completely corrects the abnormal quantitative relationship of these electrolytes. These effects are due largely but not entirely to an action upon the renal tubules (p 451). They are partly extrarenal for this steroid appears to influence membrane permeability of the tissues generally. As a result of its effect in causing the retention of sodium and water, desoxycorticosterone administration, especially if the intake of sodium is increased, may be followed by a great increase in plasma volume and edema. In overdosage, more serious effects may result, e g, hypertension, dilatation of the right ventricle and

in patients receiving large doses of desoxycorticosterone (see fig 59 6)

An influence upon growth has been demonstrated (by Hartman and Thorn) for desoxycorticosterone and the amorphous fraction. Young adrenalectomized animals show retarded growth which is resumed when either of these preparations is administered. Corticosterone and compound E, on the contrary, actually inhibit growth.

The reputed action of cortical principles in reducing capillary permeability should be mentioned since such an effect affords a possible theoretical basis for their use in surgical and burn shock. Menkin has reported that the dilatation and increased permeability of the capillary vessels caused by inflammatory exudates (see p 98) can be reduced or prevented by the administration of extracts of the adrenal cortex.

Table 80 summarizes the principle physiological actions of the different fractions of the adrenal cortex.

EFFECT OF DOCA UPON THE CONCENTRATION OF INTRACELLULAR Na AND K

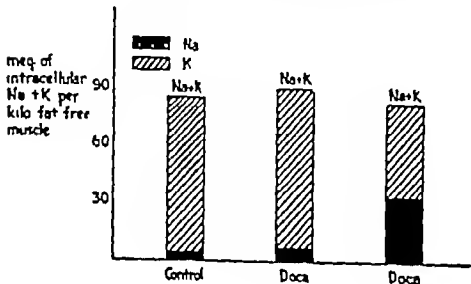


FIG 59 6 The replacement intracellularly of potassium by sodium in the muscles of normal dogs under the influence of desoxycorticosterone acetate (DOCA). The effect is inhibited by the administration of KCl (After Loeb)

pulmonary congestion. Death from cardiac failure has resulted from its clinical use. The rapid increase in plasma volume, and, as a result of this, the extra burden thrown upon the cardiovascular system are thought to be the most important factors leading to these ill-effects. Depression of the potassium concentration of the body fluids may be the lethal factor in some instances. Loeb and his associates have shown that the administration of desoxycorticosterone to animals may lower the potassium level to the point where paralysis results. In dogs given desoxycorticosterone daily for several successive days in succession, a condition resembling diabetes insipidus develops. The water intake, especially if salt is administered as well, is increased by from six to ten fold. The output of urine rises accordingly and its specific gravity falls. The administration of potassium salts or a reduction in the sodium intake is undoubtedly of benefit

TABLE 80

FRACTION OF CORTICAL EXTRACT	MAIN PHYSIOLOGICAL ACTION
Compounds with an oxygen atom at C ₁₁ , e g, corticosterone, dehydrocorticosterone and 17-hydroxy-11-dehydrocorticosterone (cortisone)	(a) Gluconeogenesis from protein, leading to hyperglycemia, and increase in the glycogen stores (b) protein breakdown, (c) mobilization of fat, (d) minimal effect on water and salt metabolism
Desoxycorticosterone	(a) Greatest effect next to electrocortin (p 838) on water and salt metabolism, (b) most effective in sustaining life in adrenalectomized animals, (c) most effective in postponing muscular fatigue, (d) second only to the amorphous fraction in restoring renal function, (e) but little effect on carbohydrate, protein or fat metabolism
Amorphous fraction	(a) Most effective in restoring renal function, (b) nearly as effective as desoxycorticosterone in sustaining life in adrenalectomized animals, (c) weak action in postponing muscular fatigue.

The action of the adrenal cortex upon lymphoid tissue and the production of plasma protein is discussed in chapter 10

Cortisone

Great interest has been aroused within recent years, and both laboratory and clinical research into the functions of the adrenal cortex have been intensified by the dramatic discovery of Hench and his associates of the therapeutic value of compound E (cortisone) in acute rheumatism and rheumatoid arthritis. In acute rheumatism there is rapid reduction of fever, and of the swelling and pain in the joints, following the oral or parenteral administration of cortisone. The high sedimentation rate which occurs in this disease is restored to the normal value by this steroid. Compound F (hydrocortisone) and the adrenotrophic hormone of the hypophysis, ACTH (through its stimulation of the production and liberation of the cortical hormones) have a similarly beneficial effect. Rheumatoid arthritis is in most cases benefited by these compounds, sometimes dramatically, but improvement as a rule is not sustained after treatment has been discontinued.

The adrenal steroids effect amelioration of these arthritic diseases though there is no evidence of adrenal insufficiency. Since such large doses are required, much greater than are employed in the treatment of Addison's disease, it is not credible that these compounds act by correcting a hormone deficiency. That they act pharmacologically, rather than by replacement, there can be little doubt. In the dosage employed they appear to exert a specific effect upon the metabolism of the collagenous tissues (connective tissue fibers embedded in cement or ground substance containing hyaluronic acid) of joints, skin, arteries and many other structures. The results of Opsahl's experiments may provide a clue to this action. She found that in mice, adrenal cortical extracts inhibited the enzyme hyaluronidase, as shown by a reduction of the "spreading effect" of the enzyme. Compounds A, E, and F exhibit this inhibiting action, but Compound E was the most effective.

The *therapeutic value* of cortisone and dehydrocortisone has been investigated in a large number of diseases, many of them apparently quite unrelated to one another. The diseases upon whose course cortisone, dehydrocortisone or ACTH has been found to have a favorable effect, though in many cases this has been slight or evanescent,

can for the most part be grouped under three main headings, as shown below

Collagenous diseases (collagenoses)

- Acute rheumatism (rheumatic fever)
- Rheumatoid arthritis
- Rheumatoid spondylitis
- Psoriatic arthritis
- Scleroderma
- Periarteritis nodosa
- Dermatomyositis

Allergic diseases or diseases due to abnormal sensitivity of cells

- Bronchial asthma
- Allergic rhinitis (hay fever)
- Contact or atopic dermatitis
- Urticaria

Lymphatic and myelogenous diseases

- Acute and chronic lymphatic leukemias
- Lymphosarcoma
- Hodgkin's disease
- Lymphoid thyroiditis (struma lymphomatosa)
- Myelogenous leukemia
- Sarcoidosis
- Multiple myeloma

In the following list are a number of other pathological states in which ACTH or cortisone has been employed with variable success and which cannot be placed in any of the above categories. Ulcerative colitis, acute inflammatory eye diseases, alcoholism with delirium tremens or Korsakoff's psychosis, chronic nephrosis, Waterhouse-Friedrichsen syndrome, Loeffler's syndrome, and the adrenogenital syndrome.

Deleterious effects which may result from cortisone or ACTH treatment. A number of serious adverse effects may result from treatment with these agents. Prolonged cortisone therapy may lead to adrenocortical atrophy. Collapse and death from acute adrenal insufficiency may follow cessation of treatment. Several instances of the development of Cushing's disease (ch 57) e.g., hirsutism, rounding of the facial contours, osteoporosis, kyphosis and hypertension have been reported from the continued administration of cortisone or ACTH. Mental aberrations ranging from an exaggerated sense of well-being (euphoria), moodiness or mild anxiety states to definite psychoses, e.g., paranoia, have been reported following cortisone therapy. Wound healing is apparently delayed by cortisone, and resistance to certain infections is lowered. Hamsters can be infected with polyomyelitis virus by intraperitoneal injection, and develop a highly virulent form of the disease if pre-treated with cortisone, but animals similarly injected, but not receiving the hormone, could not be infected (Shwartzman). The spread of a tuberculous infection from a previously localized focus is thought to be

another serious effect which may sometimes follow cortisone therapy. Peptic ulcers not uncommonly develop under treatment with these agents and perforation of the bowel in cases of ulcerative colitis so treated have sometimes occurred.

The adrenal cortex and the sex functions The following observations point to the cortex as being in some way associated functionally with the gonads. (a) Animals during heat or pregnancy or injected with the luteinizing principle of the hypophysis, withstand adrenalectomy better than at other times, and progesterone lengthens the life span of adrenalectomized rats. Bülbürg has shown a corresponding relationship between the testes and corticosterone. In drakes the seminiferous tubules atrophy between breeding seasons and birds adrenalectomized at the latter time require less corticosterone for maintenance than do those adrenalectomized during the breeding season. A direct correlation was observed between testicular size and the quantity of hormone required. The greater amount necessary for maintenance in the breeding season may be due to an antagonism between the actions of the adrenal and testicular hormones, or, as Parkes suggests, to the production of a principle with adrenal cortical activity by the testes in the interbreeding season. (b) Estrogens and androgens can be extracted from the adrenal cortex and are found in normal urine. The excretion of these substances also occurs in eunuchs and in ovariectomized women, which indicates that they are derived, in part at least, from the adrenal cortex. It is not improbable that normally the adrenal cortex, through the manufacture and liberation of these hormones plays a rôle in the control of the sex functions. The persistence of the sex libido after the excision of the gonads (p. 872) conforms with such an idea. (c) Corticosterone and the sex hormones are closely related in chemical structure. Desoxycorticosterone exhibits progesterone activity. (d) the cortex enlarges during pregnancy, and special cells (cells of Stilling) appear in the adrenal cortex of the frog during the mating season. (e) the ovaries of hypophysectomized tadpoles are increased in weight by injections of cortical extract. (f) in women the intermenstrual periods are said to be shortened by from 3 to 5 days by the administration of cortical extract (Hartman and associates). (g) cortical extracts have been reported to cause precocious sexual maturity in rats. (h) sexual abnormalities are striking features associated with tumors of cortical tissue (p. 847). (i)

growth of the gonads of both male and female rats is stimulated by feeding with adrenal tissue. (j) the origin of the genital organs and cortex from a common embryonic tissue (celomic epithelium).

The regulation of cortical hormone secretion, ACTH liberation Though the secretion of the adrenal medulla is controlled by nerves (cholinergic, ch. 72), the output of cortical hormones is regulated by the adrenocorticotrophic hormone. But since the liberation of the latter is under hypothalamic control, and the hypothalamus is influenced through impulses from higher centers of the brain (states of stress), and through the liberation of adrenaline, the cortex is under *indirect* nervous control.

Since the discovery of the eosinopenic and lymphopenic responses to ACTH, the control of the secretion of this trophic hormone has been rather thoroughly investigated. The regulating mechanisms are extraordinarily complex, and not yet fully understood, but the following picture depicting the integration of the hypothalamus, adenohypophysis, adrenal medulla and adrenocortical hormones to bring about a kind of chain reaction cannot be far from the truth. Hypothalamic function can be depressed or stimulated by impulses from higher cerebral centers as well as by injury, or some abnormality of the peripheral tissues, the impulses ascending by sympathetic and somatic afferents and fiber tracts of the spinal cord. From the hypothalamus sympathetic efferents descend to the adrenal medulla and cause the release of adrenaline. The latter, carried in the blood stream, acts in turn upon the hypothalamus and possibly also directly upon the anterior pituitary itself.^a Injections of insulin or of histamine (the former through its hypoglycemic effect) also stimulate the hypothalamus. Acted upon in any of these ways, the hypothalamus causes the release of ACTH by the adenohypophysis through its neurohumor conveyed in the hypophyseal-portal vessels, the output of adrenocortical steroids is increased, eosinopenia and lymphopenia and depletion of the cholesterol and ascorbic acid content of the cortex result. But the rise in the blood level of the adrenal steroid hormones puts a check upon the output of ACTH. We may say, then that the mechanism governing the activity of the adrenal cortex is largely self-regulated. There is some question whether nerve impulses from injured tissues can of themselves, i.e., in the absence of a

^a Noradrenaline in physiological dosage does not appear to exert this effect.

reflex liberation of adrenaline, cause ACTH release. Besides the release of ACTH through impulses initiated in injured tissues, it is possible that a product of cell destruction (histamine?) may act humorally upon the hypothalamus either directly or indirectly through the release of adrenaline.

Some of the evidence for the foregoing account of the mechanisms governing ACTH release may now be cited. The eosinopenic or lymphopenic response or the reduction in adrenocortical ascorbic acid is employed to detect the release of ACTH.

(1) Under conditions of stress ACTH is liberated if the hypophysis is intact but is abolished by complete section of the neural stalk or by a lesion in the posterior part of the tuber cinereum. (2) Electrical stimulation in the latter region by means of a buried secondary coil and a large primary coil surrounding the animal's cage (i.e., by remote control in order to obviate any extraneous stressing stimulus) causes ACTH release (see Harris). (3) ACTH liberation following tissue injury (e.g., burning) is almost completely abolished by sympathectomy and section of the spinal cord above the level of the injury. The slight eosinopenic response which is still obtainable may be due, as mentioned above, to products of tissue destruction. (4) Section of the spinal cord at the level of the 3rd thoracic segment prevents the reflex secretion of adrenaline resulting from a scalp injury, but does not block impulses from reaching the hypothalamus. Animals, after cord section at this level, do not release ACTH in response to an injury of the scalp whereas normal animals do. This experiment suggested that adrenaline liberation is of paramount importance in the ACTH release caused by tissue injury (Long).

Methods for the bio-assay of ACTH preparations
Several methods for the biological assay of corticotrophin have been adopted. (1) Determination of the quantity of hormone required to cause a 50 per cent increase in adrenal weight, when injected into 21-day old rats, this is defined as a *normal rat unit*. (2) The amount of ACTH required to induce a fall of 17 per cent in the eosinophil count of mice. (3) *Repair test* the quantity of ACTH required in a 4 day period to restore to normal the adrenal cortex of 28-day-old female rats hypophysectomized 14 days previously. (4) *Maintenance test*. A unit of ACTH as defined by this test is the amount required over a period of 15 days to maintain the pre operative adrenal weight of hypophysectomized 40-day old male rats. (5) *Ascorbic acid depletion test* the

left adrenals of male rats (120–160 g weight) are removed 27 hours after hypophysectomy. The hormone preparation is injected intravenously and the right adrenal removed one hour later. The two sets of adrenals are removed and analysed for their ascorbic acid contents. From a comparison of the ascorbic acid concentration in the right and left sets of glands the ascorbic acid depletion of the former is estimated.

DISEASE OF THE ADRENAL CORTEX IN MAN

Addison's disease

The syndrome known today as Addison's disease was first described (in 1855) by Thomas Addison and ascribed by him to tuberculous disease of the adrenals. Experimental and clinical observations since that time have fully substantiated Addison's conclusion that the disease is due to adrenal involvement. Tuberculous disease of the gland is found, however, in only a proportion of the cases. It has also been shown that deficiency of the cortex and not of the medulla is the essential cause of the disease. Its chief features, which closely resemble those seen in adrenalectomized animals, are (a) muscular weakness and languor, (b) low blood pressure and reduced circulation rate, (c) gastrointestinal disturbances, loss of appetite (anorexia), hypochlorhydria and vomiting, (d) pigmentation of the skin and mucous membranes, bronzing, tanning or a dirty brown cutaneous discoloration being a classical symptom of the disease, (e) lowered metabolic rate, subnormal temperature, sodium loss and a rise in serum potassium, reduced blood volume (plasma loss), dehydration and loss of weight, (f) renal insufficiency with consequent rise in blood nonprotein nitrogen, (g) depression of the sexual functions, (h) hypoglycemia may occur and be the immediate cause of death. (i) Abnormal electroencephalogram, slowing of alpha rhythm and reduced number of beta waves. (j) Changes in the electrocardiogram, e.g., low voltage, flat or inverted T, prolongation of P-R and QRS and depression of RS-T. The pigmentation of the skin may be so deep that the patient is mistaken for a mulatto, it is most pronounced in those regions, nipples, abdomen, etc., where the normal pigmentation is greatest (fig. 59.7). The palms of the hands and soles of the feet remain pale. The discoloration is due to the excessive accumulation of the normal cutaneous pigment, *melanin*. This is deposited chiefly in the basal cells of the epidermis, but pigment granules are also found in the dermis. The change in pigment metabolism which

causes the deposits is unknown.⁹ Szent-Györgyi found that ascorbic acid which, as already mentioned, is present in relatively high concentration in the adrenal cortex inhibits pigment formation in plant tissue but it remains to be shown that a

Addison's disease, unless treated with cortical hormone or with a diet of high salt and low potassium content, is almost invariably fatal within from 1 to 3 years. The high salt, low potassium diet exerts an almost specific effect in adrenal insuffi-



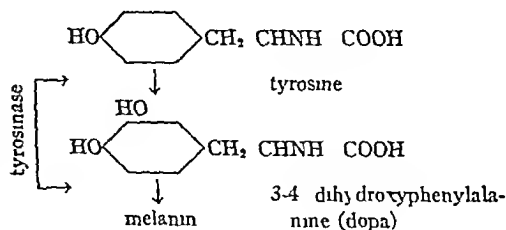
FIG 597 Showing pigmentation of the skin together with patches of depigmentation in a patient with Addison's disease (After Kepler)

lack of this substance in Addison's disease is responsible in any way for the bronzing

⁹ Bloch found that sections of normal skin became deeply pigmented when placed in a dilute solution of 3,4-dihydroxyphenylalanine, albino skin treated similarly remained unpigmented. Solutions of other aromatic compounds (e.g., tyrosine, tryptophane, pyrogallol, etc.) did not cause pigmentation of skin sections. Bloch concluded that dihydroxyphenylalanine, which he called *dopa*, was the precursor of melanin, the conversion being brought about by an oxidizing enzyme (dopa oxidase) in the skin. Until recently tyrosinase, present in potato, fungi and other plants, and which is capable of catalysing the conversion of tyrosine to "dopa", has been believed to be absent from mammalian tissues. But since its discovery in animals this enzyme has been assigned the role of catalyzing the reactions, tyrosine to "dopa", and "dopa" to melanin (see Lerner and Fitzpatrick). Tyrosinase is now believed or thought, to be identical with Bloch's dopa oxidase. See scheme below.

The reactions resulting in melanin production are inhibited by ascorbic acid, and, under certain circumstances, by cortical steroids. So far, the foregoing facts have not led to a satisfactory theory of the cause of the pigmentation in Addison's disease.

ciency, but desoxycorticosterone acetate is most usually employed, it is given in oil or as pellets inserted subcutaneously. However, as already mentioned, there are certain disadvantages associated with the use of this corticoid, a marked anti-diuresis with sodium retention and loss of potassium occurs when a certain dosage is exceeded. A high Na and low K in the diet enhances the adverse effects. Nor does the desoxy compound constitute a complete replacement therapy for



the tendency to hypoglycemia during fasting and the electroencephalographic and electrocardiographic abnormalities are not corrected. Cortisone given with DOCA provides more complete replacement therapy in Addison's disease.



FIG 598 Enlarged abdomen and precocious sexual development of a boy aged thirty months. A mass palpated in the abdomen was probably a suprarenal cortical tumor. (After Rowntree and Ball.)

Cortical tumors—adrenogenital syndrome, pseudohermaphroditism¹⁰

Tumors composed of cortical tissue which secrete an excess of androgenic compounds may arise in the adrenal itself or in aberrant collections of cortical cells (adrenal "rests") which are found in the broad ligament of the uterus, in the neighborhood of the testes or in the retroperitoneal tissue of the abdomen or pelvis. Such growths, or a diffuse hyperplasia of the adrenal cortex, may occur in children or in adults, and are associated with extraordinary abnormalities of development, especially of the *accessory organs of sex* and of the *secondary sex characters* (p. 871).

When cortical hyperplasia or a tumor composed of cortical tissue occurs in young children, puberty appears prematurely, a male child of 4 or 5 years, for example, may show the sexual development of an adult (fig. 598). The testes and penis are enlarged, hair may appear upon the chest, pubis and

¹⁰ Pseudohermaphroditism is defined as a condition in which the gonads of one sex only are present but the external genitalia combine features of both sexes, so that from inspection of them alone it is difficult or impossible to determine the sex. In true hermaphroditism both ovarian and testicular tissues are present, it is an embryological aberration and not due to a hyperfunctional adrenal cortex (see also ch. 61).



FIG 599 Virilism due to adrenal tumor. On left, at age of 28 years before the onset of the disease, on right, at age of 35 years at height of disease. (After Lescher.)

face, and there may be precocious sexual desire. Often there is also unusual muscular development or extreme obesity. There is early ossification of the epiphyses. Male children showing these characters have been described as resembling "an infant Hercules" (Herculean type of Weber), or "a burly brewer's drayman" (Guthrie). Growth is rapid as a rule, but the epiphyses fuse early, young subjects of the disease therefore do not reach full stature.

In little girls the breasts hypertrophy, hair appears on the mons veneris and around the vulva, the uterus develops prematurely, the clitoris is hypertrophied and there is sometimes precocious menstruation. In appearance such children resemble stout little women. Women who are subjects of this disease become mannish in appearance and disposition (virilism, fig. 39.9). The voice deepens, menstruation ceases, the breasts atrophy and hair may grow upon the face, chest and limbs, homosexuality is a common feature. The urinary excretion of androgens (17-ketosteroids) is increased several fold. In some adult female subjects, the virilism is accompanied by glycosuria and decreased sugar tolerance, due apparently to an increased production of glucocoids. This type is known as the Achard-Tiers syndrome and was described by these authors as the "diabetes of bearded women". Adult male subjects of cortical tumors, as a rule, give no evidence of endocrine disturbances. In some, however, an exaggeration of the masculine characters is manifest, e.g., enlargement of the penis, a tendency to hirsutism and increased sexual appetite. In others the tumor secretes excessive amounts of estrogen when a certain degree of feminization is observed, e.g., enlargement of the mammae, atrophy of the testes and a feminine distribution of fat.

It might be expected that with the increase in active adrenal cortical tissue, changes in electrolyte balance corresponding to those following injections of adrenal cortical hormone, namely, a rise in blood sodium and a fall in potassium, would be found. But actually such changes occur only in a proportion of cases and are, as a rule, moderate in degree. On the other hand there may be an associated depression of adrenal salt retaining factors with the appearance of Addisonian symptoms, due to atrophy of the opposite adrenal. Surgical removal of the tumor or hyperplastic tissue is the preferred treatment when possible. Sometimes as a result of atrophy of the normal adrenal tissue, acute cortical insufficiency develops after operation. In order to forestall such a crisis treatment

with a cortical hormone is instituted pre-operatively. The administration of cortisone is sometimes successful in suppressing adrenal hyperplasia, the elevated blood level of this hormone tends to inhibit the output of ACTH and, as a consequence, to reduce its stimulating effect upon the hyperplastic tissue. This treatment however appears to be of little benefit if the disease is due to a tumor whose activity appears to be independent of pituitary control.

Brosier and Vines have made the interesting discovery that hyperplastic adrenal cortical tissue or tumor tissue in cases of virilism contains large numbers of cells which have an affinity for fuchsin (fuchsinophil cells). Cells possessing this property are normally absent from the mature adrenal or present only in very small numbers. They are a characteristic feature, however, of the fetal adrenal of both sexes between the 10th and 17th weeks (in the case of the male) and between the 11th and 15th weeks (in the case of the female) but disappear thereafter. These observers suggest that virilism is due to the elaboration of a masculinizing hormone by the fuchsinophil cells and that the female fetus normally passes through a male phase. The importance of the fuchsinophil cells in the production of hyperadrenocortical manifestations has been questioned however by Sofer.

Clinical tests of adrenal cortical function

In mild or atypical cases of adrenocortical insufficiency the following tests may be employed for its detection.

(1) *Eschschertic response to ACTH* is one of the most sensitive tests. In normal persons a 50 per cent reduction in circulating eosinophils results from the administration of a test dose of corticotrophin, whereas the response is absent or much reduced in adrenocortical insufficiency.

(2) *Rebinder Kepler Pinner test* to ingested water. Normally diuresis commences within 30 to 45 minutes after water drinking in adrenal insufficiency the excretion of the excess water is much delayed.

(3) *Cutler-Pinner-Wilder test*. The urinary excretion of sodium chloride during a 53 hour period of rigidly restricted salt intake accompanied by the administration of potassium. Normally less than 150 mg. per 190 cc. of urine is excreted in the last 4 hours of the test. The subject of adrenal insufficiency excretes a much larger quantity. This test is potentially dangerous, for unless the adrenal insufficiency is of mild or moderate degree, an adrenal crisis may be precipitated.

Other determinations, e.g., the tendency to hypoglycemia during fasting, the sensitivity to insulin, estimation of urinary 17-KS (which are reduced in adrenal insufficiency and increased in normals by ACTH administration) and the urea acid-creatinine ratio which is normally increased by cortical stimulation.

CHAPTER 60

THE PARATHYROID GLANDS AND CALCIUM METABOLISM

THE PARATHYROID GLANDS

Development and structure

These glands were first recognized as separate structures and described by Sandstrom (1880). He gave them their present name in the belief that they were remnants of embryonic thyroid tissue. They arise, however, quite independently of the thyroid from the endodermic lining of the III and IV visceral clefts in close association with the origin of the thymus. Nor have they any known functional relationship with the thyroid. The parathyroids are usually stated to be four in number, two in relation to the dorsal surface of each thyroid lobe. The upper one on each side is called the "external" or "parathyroid IV." It lies near the upper pole of the thyroid and is often embedded in the latter's substance. The lower one—the "internal" or "parathyroid III"—is situated near the lower pole of the thyroid or at a variable distance below this level and nearer the mid-line of the neck. It is drawn into a position below (caudal to) parathyroid IV in early embryonic life. The number of parathyroids and their positions are, however, very inconstant, in man as well as in animals, and accessory parathyroid tissue may be found anywhere in the neck or even in the thorax embedded within the thymus. The Roman numerals III and IV refer to the gill cleft from which the gland arose. The terms "internal" or "external" signify the proximity of the gland to the mesial or to the lateral aspect, respectively, of the thyroid lobe. The human parathyroids are roughly oval in shape, and about 6 mm. in length.

The parathyroid tissue is composed of large round cells, closely packed into masses separated by capillary blood sinuses, the cells rarely show an alveolar arrangement, in the adult gland they are of two main types: (a) *Clear chief cells*. These have large nuclei, the protoplasm stains poorly and is usually non-granular. Such cells are, apparently, the essential secreting cells of the gland, since they are the only ones present in the human gland up to the 10th year, and at any age in some animals, e.g., the dog. (b) *Oxyphil cells*. These are larger than the preceding and contain granules in their cytoplasm which stain with eosin. They do not appear in the human gland until after the 10th year and are entirely absent from the glands of the dog.

The *blood supply* of the glands is very rich, it is derived from branches of the *inferior thyroid artery*. The *nerve supply* is scanty and the fibers are probably entirely vasomotor in character. They have the same origin as those innervating the thyroid. Secretory nerves have not been demonstrated. The functional

activity of the parathyroids is possibly controlled by a hormone liberated by the anterior lobe of the pituitary (p. 844).

THE EFFECTS OF PARATHYROIDECTOMY— HYPOPARATHYROIDISM

The small size of the parathyroids in laboratory animals and the fact that they are often embedded in the thyroid make their separate removal by operation very difficult if not impossible. The usual experimental procedure is excision of the thyroid. When this is carried out upon cats or dogs a large enough proportion of parathyroid tissue is usually removed along with the thyroid to produce urgent symptoms of parathyroid deficiency. These symptoms constitute the condition known as *tetany*. Since tetany may be produced in other ways (see below) the condition when induced by parathyroid removal is called *parathyroid tetany* or *tetania parathyreopriva*. Tetanic symptoms do not, however, invariably result from complete thyroidectomy. The failure of symptoms to appear is due most likely to the presence of accessory parathyroid tissue, or to one or more of the four glands having been left behind, since it is not unusual for a parathyroid, especially one of the lower pair, to be situated a short distance from the thyroid. A description of the tetanic state is given on p. 850.

In general, young animals are more susceptible than older ones to the effects of parathyroid deprivation. A meat diet also appears to increase the susceptibility. But apart from the factors of age and diet the severity of the tetanic manifestations and the time of their appearance after thyroparathyroidectomy vary greatly in different species and between different individuals of the same species. This variability is also in many cases due to differences in the amount of accessory tissue which exists or to the inconstancy in the positions of the glands. In the rabbit, for example, the internal (lower) parathyroids usually lie below the thyroid and unless these are sought for and removed tetany does not, as a rule, develop. Tetany is induced in the rat with difficulty by thyroparathyroidectomy, yet the condition can be induced by other means (p. 852). In cattle also this operation is not followed by tetany.

In the human subject tetany sometimes follows thyroidectomy for goiter, or for malignant disease

and is then due to the inadvertent or unavoidable removal of the parathyroids (*post-operative tetany*)¹ The removal of a parathyroid tumor is also frequently followed by tetany (p 859) Some of the earliest reports of tetany following thyroidectomy came from the surgical clinics of Kocher at Berne, and Reverdin at Geneva, in the latter half of the 19th century The tetanic symptoms were then ascribed to infection of the operation wound (Kocher) or to thyroid deficiency (Reverdin) Though the parathyroids had been described only a few years previously by Sandstrom, their functions were unknown until the experiments of the French physiologist Gley who in 1892, established the fact that their removal was the cause of the tetanic seizures which followed operations for goiter

A GENERAL DESCRIPTION OF THE TETANIC STATE

A *In animals* the tetany following parathyroid extirpation has been studied most extensively Its chief features are (a) *fall in serum calcium* from the normal level of from 10 to 12 mg per 100 cc to 6 mg or less The fall may be very abrupt, reaching the latter value in twenty-four hours, or may be delayed for from forty-eight to seventy-two hours or so (b) *Rise in inorganic phosphorus of the blood* from a normal of around 5 mg per cent to between 6 and 8 mg or more *The urinary excretion of calcium and phosphorus* is reduced When the serum calcium falls below 7 mg per cent, no calcium can be detected in the urine by Sulkowitch's test, which consists of the addition of an oxalate solution to an equal amount of urine If the urine contains calcium, this reagent causes a fine precipitate of calcium oxalate to be thrown down (c) *Rapid noisy breathing*² (100 or more per minute), *high temperature* (104° to 105°F) and *tachycardia* (d) *Salivation*, often with frothing at the mouth (e) *Fibrillary twitchings of the muscles* followed by *tonic or clonic* (jerking) *muscular contractions* The jaws are clenched and the limbs are either stiffly extended or jerk violently, the head is dorsiflexed Sometimes there are automatic swimming-like movements of the forelimbs, in dogs, or, in cats, rhythmical jerking movements

¹ It is not very unusual for mild symptoms of tetany to occur after operation but to disappear later These are probably due to injury and a temporary functional derangement of parathyroid function

² The increased pulmonary ventilation, by blowing off carbon dioxide and producing a condition of alkalosis, no doubt increases the severity of the symptoms (see tetany of hyperpnea)

of the paws, as though the animal were trying to shake water from them (f) The muscles show *increased excitability* to the galvanic current (p 851) and to mechanical stimulation Normally 6 milliamperes are required to produce a cathodal opening contraction (C O C) when the electrode is applied to the skin over the muscle or its nerve In the tetanic state less than 1 milliamperes may be effective The time constant of accommodation is increased (see p 919) (g) It has been shown by Imrie and Jenkinson that the *phosphocreatine* of the muscles is reduced and its rate of resynthesis slower than normal (h) *Death usually occurs from asphyxia*, due to spasm of the laryngeal and thoracic muscles

The tetanic symptoms are closely related to the serum calcium level As this becomes lowered the symptoms, mild at first (perhaps merely slight stiffness of the hind limbs), become gradually more severe and when the serum calcium has fallen to between 5 and 6 mg per cent, the tetanic state is usually fully developed

Parathyroidectomized rats show an increased appetite for calcium, when given a choice between a calcium solution and water they drink more of the former than do normal rats under the same circumstances As compared with normal animals they drink less of a phosphate solution

B *In man* In infants and very young children tetany is usually seen in association with rickets In adults, it is most frequently due to parathyroid deficiency—the result of an operation for goiter or for the removal of a parathyroid tumor (p 859) The symptoms are usually less intense than those seen in parathyroidectomized animals, the condition, as a rule, running a more chronic course. Rapid respirations and high temperature are not usually seen The serum calcium does not, as a rule, fall below 7 or 8 mg per cent Neuromuscular hyperexcitability is the outstanding feature Though jerking movements or generalized convulsions may occur in children they are unusual in adults The hypertonic state of the muscles causes the hands and feet to be drawn into typical attitudes which are spoken of as *carpo-pedal spasm* (fig 60 1) The hands are flexed at the wrists, and the fingers flexed at the metacarpo phalangeal but extended at the interphalangeal joints The thumb is adducted into the palm This position constitutes the so-called *accoucheur's hand* of tetany The feet are extended at the ankles and the toes plantarflexed Spasms of the eye-muscles may be

seen, and occasionally spasmodic retention of urine occurs. In infantile tetany spasm of the muscles of the glottis is not uncommon, causing inspiratory stridor (*laryngismus stridulus*). When severe, the laryngeal spasm causes complete closure of the glottis for a time, cyanosis results and when asphyxiation seems imminent a sharp in-



symptoms appear. Emotion, some undue strain upon the organism, e.g., pregnancy, lactation or a failure in general health, may, however, precipitate an attack of manifest tetany in a subject who had been suffering from the disease in latent form. Certain tests are employed to unmask this incipient form of the disease: (a) *Chvostek's sign*,—tapping over the facial nerve in front of the ear causes twitching or spasm of the facial muscles; (b) *Trousseau's sign*,—occlusion of the circulation in the arm by means of blood-pressure armlet causes the hand to assume the typical attitude. The effect is probably due to the anoxemia induced in the muscles of the hand and forearm; (c) *von Bonsdorff's phenomenon* is the facilitation of the muscular spasms caused by hyperventilation while the circulation to the arm is occluded; (d) *Erb's sign*,—increased excitability of the muscles to the galvanic current already referred to.

OTHER FORMS OF TETANY

These with the chief changes in blood chemistry are shown in table 81.

Infantile and idiopathic tetany

Tetany occasionally arises spontaneously in infants and may then be due to parathyroid deficiency. Spontaneous, or idiopathic tetany as it is sometimes called, occurs also, though rarely, in adults as a result of defective parathyroid function. Tetany arising in this way does not differ essentially from that following parathyroidectomy. The tetany of infants is usually, however, an accompaniment of rickets and, so far as is known, is then not due to parathyroid deficiency. During the active stage of rickets the serum calcium is little if at all depressed, but during the healing stage of the disease calcium is diverted to the bones and the calcium of the serum falls. It is at this time that tetany occurs. Tetany may also be produced in rachitic rats by placing them upon an antirachitic diet (Hess and associates).

The tetanies of osteomalacia and sprue

Osteomalacia is a disease of the bones of adults (p. 768). Its pathogenesis is essentially the same as that of rickets. The serum calcium is often very low, tetany is of common occurrence. Tetany occurs in *celiac rickets* and *non-tropical sprue* (Table 81). In the latter disease the absorption of fat and of calcium is defective and the serum calcium

FIG. 601 Tetany. Description in text. (Upper photograph after Purvis Stewart, lower, after Cabot.)

spirations occur accompanied by a high-pitched "crowing" sound. These various forms of muscular spasm are grouped under the general term *spasmophilia*.

Latent tetany

Frequently the serum calcium remains just above the critical level at which definite tetanic

depressed Tetany also sometimes occurs in tropical sprue.

Tetany associated with alkalosis (gastric, hyperpneic, and bicarbonate tetany)

In pyloric obstruction, dilatation of the stomach, or as a result of persistent vomiting from other causes, the loss of chloride in the vomitus causes a change in the acid-base balance toward the alkaline side Tetany follows Alkalosis is also the

dogs, it is only after the injection of the alkaline salts, however, that tetany occurs This difference is probably due to the different effects of the two salts upon the acid base balance—the one tending to cause alkalosis and a reduction in the ionization of calcium, the other acidosis and a relative increase in the concentration of calcium ions (see below) The injection of a neutral mixture of the two salts or of phosphoric acid itself does not cause tetany

TABLE 81*

Types of tetany

	CALCIUM	BICARBONATE OF BLOOD	CHLORIDE	pH	PHOSPHORUS
Infantile or idiopathic tetany	Reduced	Normal	Normal	Normal	Normal or reduced
Tetany of osteomalacia	Reduced	Normal		Normal	Normal or reduced
Tetanus of sprue and celiac rickets	Reduced	Normal			
Gastric tetany	Normal	Increased	Reduced	Increased	Increased
Bicarbonate tetany	Normal	Increased	Reduced	Increased	
Hyperpneic tetany	Normal	Increased		Increased	
Parathyroid tetany					
(a) Experimental	Reduced	Normal		Normal	Increased
(b) Post-operative	Reduced	Normal			
Phosphate tetany (Na HPO_4)	Reduced	Normal	Normal or increased	Normal or increased	Increased
Citrate tetany	Reduced	—	—	—	—
Tetany due to calcium and vitamin D deficiency	Reduced	—	—	—	—
Tetany due to magnesium deficiency	Normal	—	—	—	—
Guanidin tetany	Normal or slightly reduced	—	—	—	Increased

* With modifications and additions from MacCallum

apparent cause of the tetany which results from increased pulmonary ventilation In this case it is the excessive elimination of carbon dioxide which is the cause of the increased blood alkalinity Alkalosis is also evidently the cause of the tetany-like seizures which sometimes follow the administration of large quantities of sodium bicarbonate for therapeutic purposes

Phosphate tetany

Phosphate tetany is produced experimentally The intravenous injection of 0.5 gram per kilogram of either the acid or the alkaline sodium (or potassium) phosphate into animals causes a profound and rapid fall in the serum calcium In

Citrate tetany

An intravenous injection of sodium citrate is a less sure way to induce tetany, but in a certain proportion of animals typical convulsions follow within 15 or 20 minutes after the injection The serum calcium is lowered

Tetany due to calcium and vitamin D deficiency

The serum calcium of rats on a diet lacking in calcium, falls, after a period of from 3 to 7 weeks, to a low level, but tetany does not develop unless the diet is also devoid of vitamin D Even when both calcium and vitamin D are absent from the diet, tetany does not appear spontaneously, but typical convulsive seizures can be induced by stimulation with the galvanic current or by a sudden sound

MILK FEVER Hypocalcemia and tetany sometimes occur in cows after calving as a result of the loss of calcium in the milk. The condition is treated by inflating the udder with air, which suppresses milk secretion and causes a rise in the serum calcium, or by the intravenous injection of calcium. A similar condition is seen in sheep after lambing and is then referred to as "lambing sickness" or "ewe distemper".

Magnesium deficiency tetany

The general features of this type of tetany are indistinguishable from those due to calcium deficiency. It has been produced in rats, dogs, and young cattle by feeding diets deficient in magnesium. Calves reared upon the whole milk, which has a low magnesium content (0.01 per cent), frequently show severe tetany and may die in convulsions. The blood calcium and phosphorus are within normal limits but the magnesium is reduced to little more than half the normal value. This type of tetany is not known to occur in the human subject.

PATHOGENESIS OF TETANY

MacCallum and Voetglin discovered (1908) that the serum calcium was invariably depressed in tetania parathyreopriva and that the condition was immediately relieved by the intravenous injection of calcium. In the years following the recognition of the connection between tetany and the parathyroids, and up to the work of MacCallum and Voetglin, the tetanic state was believed to be due to some endogenous toxic metabolite (e.g., guanidine) which was destroyed in some way by the parathyroids, but accumulated in the blood and tissues after parathyroidectomy.

With the general recognition of these facts the *calcium deficiency theory* of tetany came into being. The low serum calcium found in other forms of tetany, that is, produced otherwise than by parathyroid deficiency, e.g., the tetanies of rickets, osteomalacia and sprue, also indicates that calcium deficiency is the direct cause of the neuro-muscular hyperexcitability in these conditions. Depression of the serum calcium also explains the convulsions following the injection of phosphate. It is generally accepted that the determining factor in the production of tetany is the concentration of *ionized* calcium (p. 860) in the plasma and extracellular fluids of the body, rather than the *total* calcium concentration. For example, in nephritis with a low serum protein, the total calcium of the serum may be reduced to 3 or 4 mg. per cent, yet tetany does not occur, presumably because the concentration of ionic calcium has not been re-

duced to the critical level. As shown by the experiments of Loeb upon frog muscle the sodium and potassium ions tend to increase neuromuscular excitability, the calcium and magnesium ions to depress it. The calcium concentration of the tissues themselves (muscle or brain) is not altered in tetany, the increased neuromuscular excitability would therefore appear to be due to an imbalance between the concentration of ionic calcium in the extracellular and intracellular fluids. (See also p. 854.)

It is difficult to assess the importance of hyperphosphatemia as a factor in the production of tetany, for a reciprocal relationship exists in the blood between the concentrations of calcium and phosphorus. Phosphate retention or phosphate injection causes a fall in the calcium of the serum, a rise in the concentration of the calcium of the serum, on the other hand, tends to depress the blood inorganic phosphorus. Nevertheless, tetany can result from a reduction in the serum calcium as in infantile rickets and osteomalacia with a normal concentration of blood phosphate. Hyperphosphatemia, therefore, though undoubtedly increasing the severity of the tetanic symptoms does not appear to play the primary rôle in their development.

The tetany of alkalosis cannot be explained upon the basis of calcium deficiency, since in this type the serum calcium is not significantly lowered. It is suggested, however, that the shift of the acid-base balance of the blood toward the alkaline side causes a reduction in the ionic calcium fraction without altering the concentration of the total calcium of the serum. The following equation illustrates the possible relationship between the concentrations of calcium, bicarbonate, phosphate and hydrogen ions.

$$\frac{[\text{Ca}^{++}] [\text{HCO}_3^-] [\text{HPO}_4^-]}{[\text{H}^+]} = K$$

According to this equation an increase in the concentration of the bicarbonate ions or of phosphate ions or a fall in the concentration of hydrogen ions would cause a reduction in the concentration of ionized calcium without a change in the total calcium level of the serum.

Certain observations, however, are in conformity with such a hypothesis. The beneficial effect upon tetany of the administration of an acidifying salt such as ammonium chloride, since it is not accompanied by a rise in the total calcium of the serum,

and the difference, already mentioned, between the actions of the alkaline and acid phosphates (p 852), may be explained upon such a basis

The question of the neuromuscular mechanisms responsible for the tetanic seizures has not received a decisive answer. D. N. Paton and his associates sectioned the cord in parathyroidectomized dogs and observed cessation of the clonic and tonic spasms, but the tremors and the fibrillary twitchings were abolished only by section of the peripheral nerves. These results indicated that the tonic and clonic spasms were supraspinal in origin while the finer movements were dependent upon spinal centers. West more recently concluded that the supraspinal centers were not involved since the characteristic tonic and clonic spasms persisted after section of the cord in the upper thoracic region. The integrity of the spinal reflex arcs was, however, considered to be essential for the tonic and clonic manifestations since they were abolished after section of the dorsal roots. The fibrillary movements and the increased electrical excitability of the muscles appeared to be dependent upon a peripheral mechanism. They persisted for at least 24 hours after section of both afferent and efferent nerves. In contradiction of West's conclusions, Greenberg and his colleagues state that in rats, tetanic movements of the hind limbs but not of the forelimbs are abolished by transection of the cord at the level of the 7th spinal segment. They conclude that activity of nervous centers above the spinal level is essential for the development of both the tonic and clonic seizures.

TREATMENT OF TETANY

Though the tetanic symptoms are rapidly abolished by the intravenous administration of calcium salts, the beneficial effect is of short duration. Calcium by mouth is of little value in acute tetany, but is of some value in the more chronic forms. Acidifying salts, e.g., ammonium chloride are also of benefit. A single injection of parathyroid extract will relieve the condition in a few hours and hold it in abeyance for several days. In chronic tetany (e.g., post-operative) irradiated ergosterol in the form of dihydrotachysterol or calciferol is of great value, especially when combined with a high calcium and a low phosphorus intake. According to Anderson and Lvall adjustment of the calcium and phosphorus of the diet (0.5 to 0.65 gm of phosphorus, daily) is capable alone of controlling the symptoms. Lactose or dextrin in the diet tends to reduce tetanic neuromuscular hyperexcitability, for the organic acids produced during the fermentation of these carbohydrates lower intestinal pH, and thereby increase calcium absorption.

THE ACTIONS OF PARATHYROID EXTRACT (PARATHORMONE)

The belief in the calcium-regulating function of the parathyroids, which followed naturally upon the discovery that hypocalcemia was an accompaniment of parathyroid deficiency, received spectacular confirmation in 1925. In this year Collip obtained an extract from beef parathyroids which possessed a powerfully hypercalcemic effect.³ By the intravenous or subcutaneous administration of this extract to parathyroidectomized dogs the serum calcium can be maintained at the normal level. Violent tetanic symptoms are abolished within three or four hours after the injection of 10 or 20 units, and by the daily administration of considerably smaller doses than this the animal

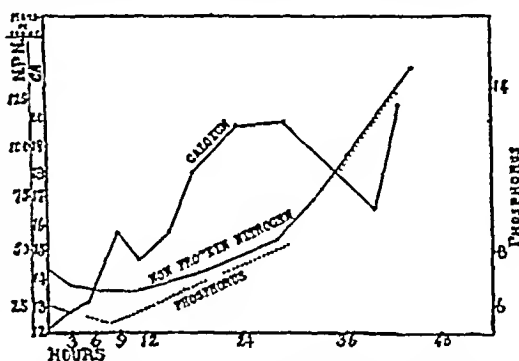


FIG 60.2 Curves of serum calcium, blood phosphorus and non protein nitrogen reproduced from Collip's article describing the effects of parathyroid extract. The curve representing the non protein nitrogen is not from the same animal as are the curves of serum calcium and blood phosphorus

is maintained in good health indefinitely. If after the calcium has been raised to normal the administration of the hormone is continued in frequently repeated doses (10 to 20 units twice daily or oftener) or if given in such dosage to a normal

³ The extract was obtained by boiling fresh glands for 1 hour with 5 per cent HCl. The resulting liquid after cooling was made alkaline to pH 8 by the addition of NaOH. Hydrochloric acid was then added slowly until a maximal isoelectric precipitation of protein occurred. The precipitate was removed by filtration or centrifuging. The filtrate or supernatant fluid contains the active principle. The potency of the extract was assayed upon dogs, a unit being defined as $\frac{1}{100}$ of the quantity of extract which will cause a rise of 5 mg per cent in the serum calcium of a 20 kilogram dog within a period of 15 hours. The extract, for which the name "parathormone" was suggested by Collip, is usually sold in 5 cc. vials and has a potency of 20 units per cc. It is effective by intravenous or subcutaneous injection, but is practically inert when given by mouth.

animal, overdosage effects are produced. These are (a) *Early changes in blood chemistry*. The serum calcium rises abruptly and within from twenty-four to forty-eight hours usually reaches a concentration of from 18 to 22 mg per 100 cc. (fig 602). During this time the inorganic phosphorus shows a moderate fall followed by a return to normal and a small rise. There is a slight rise in the potassium and magnesium of the serum. (b) *Early symptoms*. During the rise in serum calcium there are loss of appetite, depression and weakness, polyuria, vomiting, diarrhea and dehydration. (c) *The urinary excretion of calcium and phosphorus* is greatly increased. The increased excretion of phosphorus precedes the rise in serum calcium and the increase in urinary calcium. There is little change in the fecal excretion of these elements. (d) *Later blood changes* are a reduction in the hypercalcemia by 2 or 3 mg per cent, a pronounced rise in the plasma inorganic phosphorus, a four-fold increase in blood nonprotein nitrogen, a reduction in blood volume by 15 per cent due to plasma loss and, in consequence, concentration of the blood and, a great increase in its viscosity. (e) *Later symptoms*. At the time that these blood changes are occurring urgent symptoms appear,—vomiting of bloody fluid and sometimes the passage of blood-stained stools, signs of renal failure, great prostration ending in death. (f) *At autopsy* the gastro-intestinal mucosa is found to be the seat of extensive hemorrhages, and the stomach and upper part of the intestinal canal contain a quantity of bloody fluid.

All the manifestations are intensified by a high calcium diet or the administration of calcium salts.

The preceding description applies chiefly to dogs. The serum calcium of herbivorous animals, rats, mice, rabbits and guinea pigs, responds much less readily to the extract and the post-mortem picture so characteristic of its effects in dogs and cats, is not seen. In herbivorous animals, on the other hand, repeated doses cause the deposit of calcium in the soft tissue (metastatic calcification), particularly of the arterial tree, this is infrequent in dogs. In the human subject hypercalcemia is produced about as easily as it is in dogs. In both man and dogs tolerance to the hormone not infrequently becomes established after a certain number of doses.

Therapeutic (physiological) doses administered to a normal man, or to a subject of hypoparathyroidism, cause a moderate increase in P and

Ca excretion, a fall in serum P and a rise in serum Ca, i.e., a complete reversal of the biochemical abnormalities caused by parathyroidectomy (fig 603).

The excess calcium in the serum following parathormone overdosage is derived from the skeleton. Bauer, Aub and Albright have shown that repeated doses of the extract to rabbits causes a reduction of the trabeculae of the epiphyses indicating that they serve as a store of calcium, which is rapidly mobilized by the hormone (fig

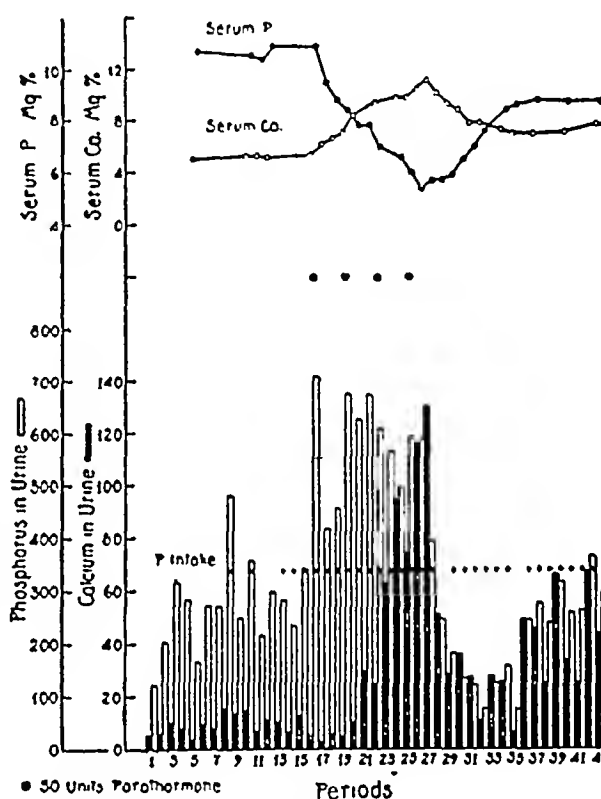


FIG 60.3 Showing the effects of the parathyroid hormone upon the serum calcium and phosphorus and upon the urinary excretion of these minerals in a case of hypoparathyroidism (After Albright and Ellsworth)

604) By the administration of parathormone over long periods a condition corresponding to osteitis fibrosa cystica of man (see below) has been produced by Bodansky and Jaffe in guinea pigs, and by others in rats and puppies. Continued administration of the hormone to experimental animals results in gross and microscopical changes in the glands themselves, they become much reduced in size and their cells appear shrunken, show hydropic degeneration and a diminution in the number of mitotic figures. Such changes indi-

cate that the administered hormone has depressed the functional activity of the glands

The fundamental nature of the hormone's action has been the subject of a considerable amount of discussion and no certain conclusion can be drawn. It has been thought by some (Cameron and Moorhouse) that the hormone controlled the formation of a specific non-diffusible organic calcium compound. The calcium level in the blood was conceived to depend upon a "series of interlocked equilibria between this compound and inorganic calcium ions." Increase in the concentration of the organic compound in the blood caused a corresponding increase in ionic calcium which was furnished by the bones (see also p 863). After parathyroidectomy, conditions were reversed, a fall in concentration of the compound and reduction in ionic calcium resulted. Another view (Greenwald) was that the hormone itself, or a substance formed through its

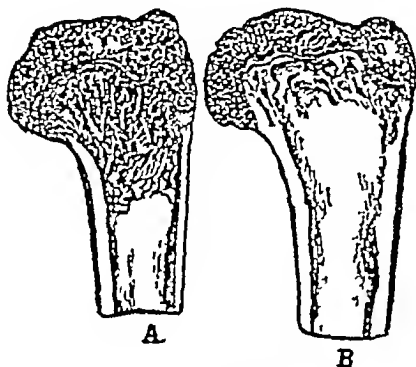


FIG. 604 Showing trabeculae of humeri of cat on (A) a high calcium diet, and (B) on a low calcium diet. (Redrawn from Bauer, Aub and Albright.)

action, increased the solvent power of the plasma for calcium. This substance, designated X, was supposed to unite with calcium ions to form an undissociated organic calcium compound which was stated to resemble calcium citrate. Through such action of the hormone concentration of calcium ions in the plasma was reduced and the concentration of undissociated calcium increased. Since the plasma was believed to be in equilibrium with solid tricalcium phosphate ($\text{Ca}_3(\text{PO}_4)_2$) of osseous tissue, a reduction in the calcium ion concentration of the plasma resulted in the liberation of calcium ions from the bones. The result was a rise in the total calcium (dissociated plus non-dissociated) of the plasma. Little direct experimental evidence can be cited in support of either of the foregoing views. Both picture the liberation of calcium from the bones as a physicochemical process, namely, that the parathyroid secretion acts not directly upon bone metabolism but by altering the electrolyte equilibria of the body fluids.

The histological studies of Selye in rats, on the other hand, suggest that the calcium is mobilized by cellular activity. During the hypercalcemia the bony trabeculae showed large numbers of osteoclasts which are believed to be responsible for the removal of calcium. When the animal became tolerant to the hormone the osteoclasts were replaced by osteoblasts and the bone density increased, i.e., calcium was deposited.

Albright and his associates adopt the view that the parathyroid hormone exerts its primary action upon phosphorus metabolism, and a secondary action upon bony structure. Ellsworth suggests the essential and specific action is upon the kidney, presumably by reducing reabsorption from the tubules as demonstrated by Harrison and Harrison. In the opinion of Albright and Reifenstein the specific action is upon the phosphorus of the tissue fluids, the hormone acting in some way upon the mineral to render it more readily excreted (or less readily reabsorbed) by the renal tubules. This is supposed to lead to the following train of events: there is first a lowering of the serum phosphorus which is followed by the release of the calcium-phosphorus salt from the bones. The resorption of bone causes in turn an elevation of the serum calcium, but no rise in blood phosphorus, this remains depressed owing to the increased urinary excretion. The product of the serum concentrations of calcium and phosphorus in milligrams remains at about the normal value of 40. The hypercalcemia is followed by the increased excretion of calcium in the urine, which tends to lower again the concentration of calcium in the blood and body fluids. This is countered and the new calcium-phosphorus equilibrium which has been established is sustained by the continued solution of bone salt. An observation reported by some workers and which has a bearing upon this subject, is the failure of the hormone to cause hypercalcemia in nephrectomized animals. However, others have not found that parathyroid extract is ineffective after removal of the kidneys (Ellsworth and Fuller, Haast and Taylor). But in either event, that is, whether or not hypercalcemia occurs after nephrectomy, osteoclastic activity is stimulated by the hormone, and resorption of bone thus induced. Furthermore, nephrectomy (which alone has no appreciable effect upon the serum calcium) followed by parathyroidectomy, causes hypocalcemia. The hormone, therefore, must have been exerting its usual effect though kidney function had been abolished.

Strong evidence for a direct action of the parathyroid hormone on bone is afforded by the experiments of Barnicot upon rats. When the parathyroid glands were removed and attached to excised pieces of parietal bone and the two together grafted to the cerebral hemisphere of a litter mate, intense osteoclastic activity and resorption of bone in the neighborhood of the parathyroid tissue were observed. Actual perforations of the bone were produced. Calciferol, and vitamin A both of which are known to affect bone

metabolism, had a local effect similar to that of the parathyroid tissue. Cholesterol, estradiol, or inert material, e.g., glass beads, were quite ineffective.

From the foregoing review of the experimental evidence it is probable that the hormone plays a dual rôle, namely, directly upon bone, probably through stimulating osteoclastic activity, as well as in some obscure way upon the phosphorus of the body fluids.

Secretion of the parathyroid hormone

The liberation of the parathyroid hormone is independent of nervous control. There is suggestive but no decisive evidence that the output of the hormone is regulated by the anterior lobe of the pituitary through a parathyrotrophic hormone (ch 57). A low level of serum calcium stimulates the liberation of the hormone by the glands, for when the thyroid-parathyroid apparatus was perfused with decalcified blood and the perfusate injected intravenously into normal dogs, their serum calcium was raised from 1.3 to 4.9 per cent within 3 hours or less (Patt and Luckhardt).

No functional interrelationship between the thyroid, adrenals or gonads and the parathyroids has been demonstrated.

Therapeutic uses of parathyroid extract

In *post-operative tetany* parathormone is of the greatest value in relieving urgent symptoms. It should be combined with large doses of calcium (40 to 60 grains daily of the chloride or larger doses of the lactate or gluconate) and with foods possessing a high calcium content, e.g., milk. Parathormone is generally considered unsuitable for prolonged administration, since it depletes the bones. Also, after a time, tolerance to its action frequently occurs. It would seem, however, to be a logical procedure to give the hormone in doses that would be just sufficient to replace the natural hormone which is lacking, together with large doses of calcium. Calcium chloride, since it supplies calcium directly, and through its acidifying action increases the ionic calcium, is of great value. It may be combined with irradiated ergosterol.

In *lead poisoning* the metal is deposited in the bones as tertiary lead phosphate ($\text{Pb}_3(\text{PO}_4)_2$) displacing calcium from tricalcium phosphate ($\text{Ca}_3(\text{PO}_4)_2$). Parathormone mobilizes the lead from the bones and increases its excretion in the urine. Ammonium chloride and other acidifying agents have a similar effect. These salts are administered combined with a low calcium diet, which also encourages the mobilization of lead.

According to Aub and his associates the increase in hydrogen ion concentration acts by converting the insoluble tertiary lead phosphate into the soluble di-lead salt. Caution should be exercised in the use of de-leading agents since serious effects may result from the sudden entrance of large quantities of the metal into the circulation. Toxic effects due to this cause are combated by a high calcium diet, which favors retention in the bones of the lead compound. When it is considered that the deposition in the bones of ingested lead is a device which protects the body from the toxic effects of the metal it is questionable whether it is desirable, as a rule, to employ de-leading agents. To remove the source of poisoning and permit the metal to be liberated spontaneously and gradually from the skeleton should be sufficient.

It has also been shown that radium deposits in bone may be liberated by parathormone and a low calcium diet.

Attempts to remove extra-skeletal deposits of calcium by means of parathormone have not met with success. This is not unexpected since the action of the hormone in mobilizing calcium and phosphorus is confined to the skeleton and may actually induce calcification of soft tissues (see p. 869).

The relation of vitamin D to parathyroid function

There are many points of similarity between the actions of *excessive dosage* of vitamin D (calciferol) or dihydrotachysterol (AT-10) and the parathyroid glands. The former, like parathormone, causes a high degree of hypercalcemia, and Hess, Weinstock and Rivkin found that in monkeys hypercalcemia is less readily induced by irradiated ergosterol after parathyroidectomy. Higgins and Sheard found that the parathyroids of chicks deprived of ultraviolet light became hyperplastic, but were restored to normal appearance by the administration of cod-liver oil. Taylor, Weld, Branion and Kay found that the toxic overdosage effects of irradiated ergosterol were less severe than usual in dogs in which an operation for the complete removal of the parathyroid tissue of the neck had been performed. They also showed that the overdosage effects of parathormone and of irradiated ergosterol were similar. Both substances cause the same degree of hypercalcemia (fig. 60.5) and hyperphosphatemia, and a rise in the non-protein nitrogen of the blood. In large doses either causes the withdrawal of calcium from the bones and increases the excretion of calcium and phosphorus in the urine. Both agents lower the renal threshold for calcium, by reducing its tubular reabsorption. The symptoms during life and the

post-mortem findings after poisoning with either material are also identical, and those species (herbivora) resistant to parathormone are similarly resistant to overdosage with irradiated ergosterol. It has been shown by others that either parathormone or irradiated ergosterol gives rise to metastatic calcification and to bony changes analogous to osteitis fibrosa cystica. Irradiated ergosterol, however, takes longer to show its effect upon the serum calcium than does parathyroid

Though others had previously described the condition it is very generally referred to as von Recklinghausen's disease. The morbid changes in the bones are, decalcification, the formation of cyst-like cavities and resorption of the bony tissue of the trabeculae and shaft which become largely replaced by fibrous tissue (fig 60 6) Histologically a great increase in the number of osteoclastic elements (p 856) is seen. The condition was shown by Mandl in 1926 to be due to an adenoma of a

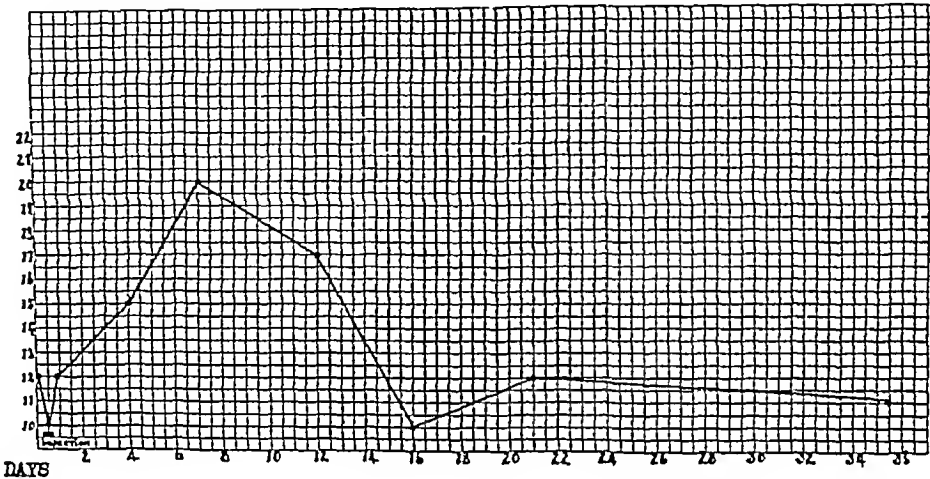


FIG 60.5 Serum calcium curve of a dog illustrating the effect of a large dose of irradiated ergosterol given intravenously, in divided doses over a period of four hours. Note the prolonged effect upon the serum calcium (After Taylor, Weld, Branion and Kay)

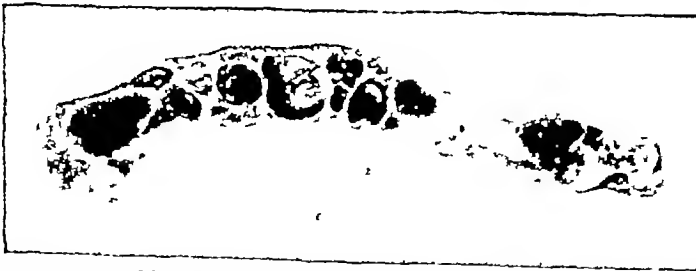


FIG 60.6 Showing section of humerus from case of generalized osteitis fibrosa cystica. (From Hunter and Turnbull after Hill and Lucey)

extract, and the hypercalcemia once established persists for some weeks (fig 60.5)

Another dissimilarity between the actions of vitamin D and the parathyroid hormone is that the vitamin, but not the hormone, increases the intestinal absorption of calcium (p 764)

Hyperparathyroidism in man. Osteitis fibrosa generalisata (osteitis fibrosa cystica)

This is a condition of the bones which was described many years ago by von Recklinghausen

parathyroid gland. Its chief clinical features are (1) Pain in the bones (2) Extreme hypotonicity of the skeletal muscles (3) Elevation of serum calcium (sometimes up to 20 mg per cent or so, but usually not above 15 or 16 mg) (fig 60.7), fall in plasma inorganic phosphorus (between 1 and 2 mg per cent) and high plasma phosphatase (4) Increase in urinary calcium, polyuria, high incidence of renal calculi, renal damage, e.g., peritubular calcium deposits leading to interstitial fibrosis, and cystic dilatation of the tubules (An-

derson) (5) There may be mental changes. The calcification appears to be secondary to a specific degenerative lesion, but the mineral deposits in their turn induce further damage of renal tissue. Deposits of calcium may occur in other tissues as well (metastatic calcification, p 869) (6) Spontaneous fractures, deformity of the bones of limbs or spine, reduced and irregular density of the bones are evident under the X-ray (7) A small tumor (parathyroid) may be palpable in the neck. In a case reported by Hunter a parathyroid tumor was found at operation behind the esophagus at the level of the second thoracic vertebra. In some instances the condition is due not to a single adenomatous tumor but to diffuse hyperplasia of parathyroid tissue.

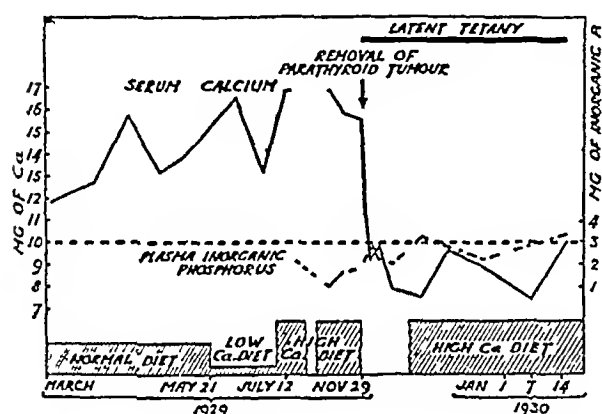


FIG 607 Showing chemistry of blood in hyperparathyroidism, and the effect of removal of the parathyroid tumor (After Hunter)

The treatment of the condition is removal of the tumor. The serum calcium falls abruptly after operation and frequently reaches a subnormal level, tetany of an intractable nature not uncommonly results. The cause of the abnormally low serum calcium is not obvious at first sight since it might be expected that the remaining healthy parathyroid tissue would be sufficient to maintain the serum calcium at the normal level. The phenomenon is probably due to the persistence after operation of a compensatory depression of activity in the other parathyroids which had resulted from the excess hormone discharged by the adenomatous gland. Bodansky and Jaffe found, for example, that in animals, hypocalcemia followed the termination of a prolonged course of parathormone injections. Depression of function, or even atrophy of an endocrine by a high blood level of its hormone is a general principle of hormone physiology. Adrenal atrophy caused by corti-

sone, and the depression of thyroid function by a high level of thyroid hormone in the blood are other examples. In the latter two instances the hypofunction is attributed to the failure of adequate amounts of the adrenocorticotrophic and thyrotrophic hormones, respectively, to be released from the pituitary.

Parathyroid hyperplasia may be induced experimentally by the following means (a) low calcium diet (rats and rabbits), (b) lack of vitamin D (chicks), (c) daily injections of phosphate over a period of weeks, (d) surgical reduction of renal tissue, or experimental nephritis, the former procedure produces dwarfing in rats and a condition resembling the renal rickets of children (Pappenheimer). In the human subject hyperplasia of parathyroid tissue may occur in association with chronic nephritis, in which event the parathyroid enlargement may be a compensatory reaction to the hypocalcemia which is the result in turn of phosphate retention and hyperphosphatemia.

In rare instances, acute hyperparathyroidism is seen with manifestations similar to those caused by overdosage with parathyroid extract. Such a state is aggravated by a high calcium intake which may so intensify the symptoms as to cause death.

CALCIUM METABOLISM

Calcium is an indispensable mineral, it is a constituent of all animal fluids and solid tissues, it plays an important rôle in a number of physiological processes and conditions. The most obvious of these are (a) coagulation of the blood (ch 12), (b) formation of bone (p 863), (c) cardiac rhythmicity (p 192), (d) maintenance of normal neuromuscular excitability (p 853), (e) milk production (p 900), (f) membrane permeability (p 31).

THE DISTRIBUTION OF CALCIUM IN THE BODY

Calcium constitutes about 2 per cent of the weight of the adult body and about 99 per cent of the total quantity is contained in the skeleton. The muscles contain about 8 mg per 100 grams of wet weight, plasma or serum from 9 to 11.5 mg per cent. The red corpuscles contain only minute amounts, the content of the whole blood is, therefore, between 4.5 and 6 mg per cent. The other body fluids, e.g., lymph, aqueous humor, ascitic and edema fluids, etc., contain it in somewhat lower concentration, while the concentration in the cerebro-spinal fluid is only about 5 mg per

cent Negligible amounts of calcium are deposited in the skeleton before the fifth month of intra-uterine life, and nearly 70 per cent of the skeletal calcium of the new-born is the result of deposition during the last two months of prenatal life (fig 608) The mother suffers a much greater loss of calcium to the suckling child Whereas, only about 20 grams of the element are lost during pregnancy, over 80 grams are secreted in the milk during a normal lactation period

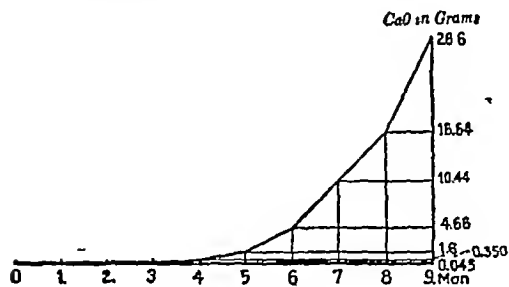


FIG 608 Showing the increase in the calcium of human fetus in later months of gestation (From Hess after Schmitz)

THE STATE OF THE CALCIUM IN THE BLOOD

Practically all the calcium of blood, as just mentioned, is contained in the plasma, after clotting it is all present in the serum, and it is upon this that determinations are usually made The clot itself contains mere traces

The calcium of blood exists in two main forms, *non-diffusible* and *diffusible* The non-diffusible form is bound to the serum proteins, both albumin and globulin fractions (but chiefly to the albumin fraction) in the proportion of about 0.84 mg per gram of protein It remains in the serum when the latter is dialyzed or subjected to ultra-filtration The non-diffusible part constitutes about 45 per cent (from 4 to 5 mg per 100 cc) of the total serum calcium The diffusible portion amounts to 55 per cent or so (5 to 6.5 mg per 100 cc) of the total calcium, when the serum is ultra-filtered this fraction passes into the filtrate *Nearly all the diffusible calcium of the serum is in ionized form* The previous low figures which have been given for the concentration of ionized calcium (1.5 to 2 mg per 100 cc) are, according to McLean and Hastings erroneous They state that only a very small amount of diffusible un-ionized calcium (about 0.25 mg per 100 cc.) is present, it is in the form of a citrate-like compound The ionized calcium is in the form

of calcium carbonate and phosphate The following is a summary of the calcium fractions in serum

Serum calcium	mg per 100 cc
Non-diffusible	4.0 - 5.0
Diffusible	5.0 - 6.5
Ionized	4.75 - 6.25
Unionized	0.25

The non-diffusible calcium, as might be expected, varies with the protein concentration of the plasma For example, in Bright's disease (with low plasma protein, and without phosphate retention) the decline in serum calcium roughly parallels the fall in plasma albumin and the hypocalcemia is due mainly to a reduction in the non-diffusible fraction Lymph, which has a lower concentration of protein than plasma has also a lower calcium content

The calcium of the cerebrospinal fluid, which is practically protein-free is almost entirely in the diffusible form, and has a concentration approximately equal to that of the diffusible fraction in the plasma The calcium concentration of the cerebrospinal fluid has, therefore, been taken as an index of the diffusible fraction of the plasma calcium upon the assumption that the former fluid is simply an ultra-filtrate This assumption is not warranted, since it is more probable that the cerebrospinal fluid is *secreted* by the choroid plexus (ch 71) Also, the administration of parathyroid extract or the injection of calcium salts, both of which raise the diffusible calcium of the plasma, causes little increase in the calcium concentration of the cerebrospinal fluid, it is little affected by parathyroidectomy

McLean and Hastings have devised a biological method for the determination of the ionized calcium of body fluids based upon the sensitivity of the frog's heart to the calcium ion concentration In this method the amplitude of the contraction of the isolated heart is recorded upon a moving drum A cannula (a modified Straub cannula) is passed through the aorta into the ventricle and tied in position A series of calcium chloride solutions graded in concentration by 0.1 millimol per liter is made up The cannula is filled with the unknown fluid (serum, edema fluid, cerebrospinal fluid, etc) and the amplitude of the heart's contraction recorded The cannula is then emptied and refilled with one of the standard solutions If the amplitude of contraction given with the latter is greater than that given by the

unknown solution, a standard solution of lower concentration is tried, if the contraction amplitude is less, the cannula is filled with a more concentrated solution. The calcium solution giving a contraction which just matches that given with the unknown fluid is taken as having the same concentration as the unknown.

McLean and Hastings state that the ionization of calcium in the body fluids is determined primarily by an equilibrium between calcium and protein which may be expressed by the following equation

$$\frac{(\text{Ca}^{++}) \times (\text{Prot}^-)}{(\text{CaProt})} = K = 10^{-2.2} \text{ (at } 25^\circ\text{C and pH } 7.35)$$

In other words, they consider that the calcium in protein-containing fluids is present as calcium proteinate which ionizes as a weak electrolyte into calcium and protein ions, with a residue of the protein-bound calcium, i.e., of the non-diffusible calcium fraction. They state that, knowing the protein and total calcium concentrations, the calcium ion concentration in human serum or other protein containing body fluid may be calculated from this equation

THE ABSORPTION AND EXCRETION OF CALCIUM— CALCIUM BALANCE

Calcium is found in food as both organic and inorganic compounds, but probably it is absorbed only in the inorganic form. Absorption occurs mainly from the upper part of the small intestine. The reaction of the intestinal contents is an important condition in the absorption of this mineral, its salts, for the most part being readily soluble in acid but insoluble in alkaline media. Gastric acidity, therefore, tends to reduce calcium absorption. Sugars, especially lactose, which in their fermentation yield organic acids in the intestine, favour absorption. Fats (free from vitamin D) reduce calcium absorption on a high Ca, low P intake, owing most probably to the formation of insoluble calcium soaps, but for some reason, perhaps the production of soluble complexes with fatty acids, such fats increase the absorption of calcium on a diet with a low Ca/P ratio. Protein food tends to increase the absorption of calcium since the latter forms soluble complexes with certain amino acids. Soluble calcium salts, such as the chloride, carbonate, lactate, and gluconate, but not the relatively insoluble phosphate, are readily and, in moderate dosage, almost completely absorbed. After the ingestion of a large dose of a soluble calcium salt, the serum calcium level rises, reaching its maxi-

mum value in about two hours. The normal serum concentration is reached again about three hours later. It is not possible to maintain the calcium level above normal for any considerable time by the administration of calcium salts.

Milk is the best dietary source of calcium, but important amounts (up to 0.2 gram daily) of available calcium may be obtained from "hard" drinking water. The calcium of many vegetables is well utilized by the rat, the calcium of carrots (and probably of certain other vegetables as well) is readily absorbed by the human intestine, being nearly as valuable as that in milk. Spinach and other plant foods containing oxalic or benzoic acid, which form relatively insoluble compounds with calcium, reduce calcium absorption. In cereals (wheat, oatmeal), owing to their content of phytic acid (inositolhexaphosphoric acid) which combines with calcium and magnesium to form an insoluble salt, much of the mineral is unavailable. The action of phytic acid in the intestine depresses the absorption of calcium in other foods as well. While wheat flour has a higher phytic acid content than has white flour and, therefore, is a poorer source of absorbable calcium. As mentioned elsewhere (ch 55), the phosphorus of phytic acid is also largely unavailable. These facts explain the decalcifying action of certain cereals.

Calcium is secreted into the small intestine in various digestive juices, but little is eliminated through the wall of the colon. The excretion of calcium continues upon a calcium-free diet or during a fast and, under these conditions, the body is in negative calcium balance. In man, on an ordinary mixed diet the calcium of the feces amounts daily to from 0.4 to 0.8 gram, this, though a considerable amount is endogenous, is mainly the unabsorbable calcium of the food.

Smaller quantities of calcium are excreted in the urine, an average of 150 mg being lost daily by this route, though there are rather wide variations. An increase or decrease in the absorption of calcium is reflected in parallel changes in the urinary excretion.

The *calcium balance*, that is the difference between the quantity of calcium ingested and that excreted in the urine and the feces, is *positive* (calcium retention) during *growth*, *pregnancy*, *acromegaly*, or after a period of *calcium starvation*. Sherman and Hawley found that children from three to thirteen years of age, upon a daily calcium intake of from 0.74 to 1.02 gram of calcium,

utilized (i.e., retained) from 0.15 to 0.62 gram per day, the quantity retained was in proportion to the size of the child (0.01 gram daily per kilogram). In adults, Breiter and his colleagues found that the utilization of the calcium of milk varied from 15.3 to 30.3 per cent. The calcium balance is negative in *infantile rickets*, *celiac* and *renal rickets*, *sprue*, *osteomalacia*, *hyperparathyroidism*, *hyperthyroidism* (pp. 858, 810), during *starvation* or *calcium deficiency*, and usually during *lactation*. In *infantile rickets* (p. 764), *celiac rickets*, and *osteomalacia* (p. 768), vitamin D administration reduces the negative balance, establishes calcium equilibrium or induces a positive balance. This vitamin increases the absorption of calcium from the intestine in normal as well as in rachitic animals and augments the deposition of calcium and phosphorus (as bone salt) in the normal grow-



FIG. 60.9 Front and side views of skeletons of twin brothers (albino rats), one of which had received a diet of normal calcium content (wheat, meat, and milk) while the other had received a low calcium diet (wheat and meat). (After Sherman and MacLeod.)

ing skeleton and in rickets. Calcium deposition is also influenced by vitamins A and C (p. 747 and p. 759). The daily calcium requirement and the calcium contents of various foods are given in chapter 56, pp. 779 and 780 (see also fig. 60.9).

BONE

The composition of bone. Osseous tissue freed from fatty marrow is composed of organic material (mainly protein), water and minerals. The chief protein constituent is *ossein*, but there are also small quantities of *osseomucoid* and an *albuminoid*. Water constitutes about 25 per cent of the bone weight, organic material 30 per cent and inorganic constituents 45 per cent. The minerals consist of Ca, P, Mg, and small quantities of potassium, sodium, chlorine, fluorine and iron. Citrate is present in bone, indeed, 70 per cent of the body's entire store of citrate is contained in the skeleton.

The existence of a diffusible citrate-like compound of calcium in serum has been mentioned. That citrate and calcium metabolism are intimately associated seems established, but the nature of the relationship is unknown (see also p. 870).

Calcium makes up from 15 to 18 per cent of the weight of fresh osseous tissue and from 20 to 25 per cent of the weight of bone which has been dried and extracted with ether.

Bone calcium exists in two forms, calcium carbonate— CaCO_3 —and tricalcium phosphate— $\text{Ca}_3(\text{PO}_4)_2$. The ash of bone amounts to about 60 per cent of its dry weight. Calcium constitutes about 36 per cent of the ash, phosphorus about 16 per cent, magnesium 0.5 per cent and CO_2 5.5 per cent. The ratio of calcium to phosphorus is approximately 2.2 to 1. The ratio of *residual calcium* (i.e., calcium other than that present as carbonate) to residual phosphorus (phosphorus not combined with magnesium) is about 2 to 1. This is, approximately, the ratio of the two minerals in tricalcium phosphate. Magnesium is present mainly as $\text{Mg}_3(\text{PO}_4)_2$. The proportions of the three chief compounds in bone ash are $\text{Ca}_3(\text{PO}_4)_2$ 80 per cent, CaCO_3 13 per cent and $\text{Mg}_3(\text{PO}_4)_2$ 2 per cent. Some of the foregoing data are given in table 82.

Rachitic bone contains a lower percentage of ash and larger proportions of water and organic material. The ratio of calcium to phosphorus, however, remains unchanged whether the rickets (experimental) develops upon a low calcium or a low phosphorus diet. The magnesium content of bone is said to be increased in rickets and in osteomalacia.

It is now generally believed that the calcium carbonate and calcium phosphate of bone are present not as separate compounds simply mixed together with smaller amounts of other mineral salts, but as a complex chemical structure.

From a comparison of the refractive indices and X-ray diffraction patterns of bone and dental enamel on the one hand, and certain crystalline minerals composed mainly of calcium and phosphorus, Taylor and Sheard concluded that the inorganic solid phase of bone and of dental enamel resembled the apatite series (podolite, fluorapatite, etc.). Fluorapatite which many consider to be nearest in chemical structure to the bone salt has the formula, $\text{Ca}_{10}(\text{OH})(\text{PO}_4)_6$. In bone, *hydroxy-*

⁴Small amounts of calcium chloride and calcium fluoride are also present.

apatite, $\text{Ca}_{10}(\text{OH})_2(\text{PO}_4)_6$ in which OH replaces F_2 , or a closely similar compound is very generally, though not unanimously, accepted as the predominant compound in bone. The bone salt is laid down in the form of minute crystals. Compounds in which sodium, magnesium and other minerals are substituted for Ca are also found in bone.

The ratio of total calcium to phosphorus, 2 2 1 (or 10 atoms of Ca to 6 of P) and that of residual calcium to residual phosphorus, 2 1 (9 atoms Ca to 6 of P) which exist in bone also support this conclusion.

Pathological calcifications, e g, salivary calculi, arterial or pulmonary calcifications, are believed to be similar in chemical composition to bone.

TABLE 82
Calcium of bone (dog)

	IN WHOLE BONE		IN BONE ASH	C P RATIO
	Fresh	Dry, ether extracted		
Total calcium (per cent)	18	25	36	2 2 1
Residual calcium (per cent)			31	2 1

Hastings and his associates hold a similar view in respect to a crystalline compound related to the apatite series as being the main mineral constituent of bone.

It is perhaps well to point out that adult bone is not simply an inert structural material but living tissue whose mineral composition fluctuates under the influence of other body functions. The trabeculae of the bone, as shown by Aub and associates, constitute a calcium store readily available when necessary for the maintenance of the calcium requirements of other tissues when the exogenous supplies are deficient. For this reason, though calcium continues to be excreted when an animal is kept, even for a long period upon a calcium-free diet, no change in the serum calcium level occurs. Parathyroid extract, as mentioned elsewhere, raises the serum calcium through its action in mobilizing these calcium stores.

In birds, the extraordinary demand made upon calcium metabolism for the production of the egg-shell has been provided for by the development of osseous tissue within the marrow spaces of the long bones. Weakening of the essential skeletal structure which otherwise might result from the

withdrawal of the mineral during the egg-laying season is thus avoided. The growth of this so-called *medullary bone* is stimulated by estrogens. Normally, its growth appears to be under the control of the ovarian secretion, it is cyclical or seasonal in character, and is associated with other phenomena of the mating season of birds, e g, the appearance of serum vitellin in the circulation, an increase in plasma fat and phospholipid, and pronounced hypercalcemia.

The bones also serve a detoxicating function, elements such as lead, radium, fluorine and arsenic, being removed from circulation and deposited in the bones and teeth. "Mottled enamel" (chalky white patches upon the surfaces of the teeth) is attributed to an excess of fluorine in the food or drinking water, though smaller amounts of fluorine are said to be beneficial for the development of the teeth and to prevent dental caries. The mobilization of bone calcium plays a very minor rôle in maintaining the normal blood reaction against the ingestion or production of excess acid. Some reduction in bone calcium can be detected, however, in animals after the administration of hydrochloric acid.

Bone formation. THE HISTOLOGY OF DEVELOPING BONE. There are two types of ossification, *intramembranous* and *intracartilaginous* or *endochondral*. The bones of the cranial vault and the mandible are formed through the ossification of membranes. The bones of the limbs and trunk and the base of the skull are first modelled in cartilage which becomes transformed into bone by both endochondral and intramembranous (i e, periosteal) forms of ossification. Studies with radioactive isotopes have shown that calcium and phosphorus are first laid down in the *epiphysis* of the developing bone, and are later transferred to the shaft (*diaphysis*) where a *primary center* of ossification appears, and, enlarging, spreads toward both ends of the bone. Secondary centers appear subsequently in the epiphyses. The deposition of calcium is hastened by vitamin D (fig 60 10). The cartilage cells just ahead of the spreading zone of calcification show active proliferation, becoming arranged in longitudinal rows. The transformation of the calcified cartilage into true bone is brought about after the following fashion. The cells of the deeper layers of the membrane covering the cartilage—perichondrium—give off long processes to form a meshwork of interlacing fibers. These cells are referred to as *osteoblasts*. The fibrous framework thus laid down soon becomes impregnated with calcium salts with the formation of a layer of true bone just beneath the perichondrium, or periosteum as it must now be called. The subperiosteal process, which is essentially

the same as that whereby the cranial bones are developed from membrane, is well advanced while the interior of the bone still consists merely of calcified cartilage. The latter, however, soon becomes invaded by blood vessels from the periosteum and by large multinucleated cells (20 to 40 μ) known as *osteoclasts*. These cells, which have a pronounced eroding action upon the mineralized cartilage, probably through the production of an enzyme, tunnel channels through it

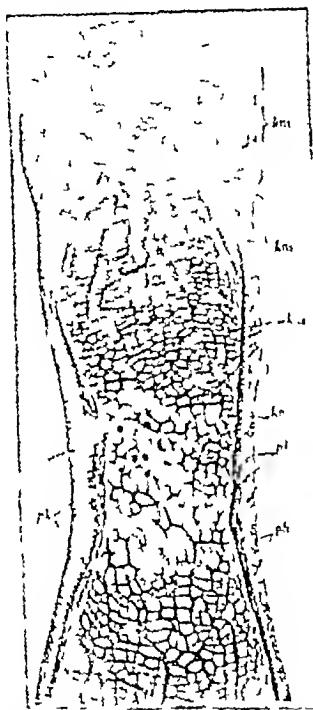


FIG. 6010. Developing bone, proximal phalanx of a three month human fetus. *km*, unchanged cartilage, *kn*, columns of cartilage cells, *km*, zone of calcifying cartilage, *km*, vestiges of walls of broken down cartilage cavities, *pk*, perichondral (periosteal) bone (black), *pl*, perichondrium (periosteum), *pk*, connection between periosteum and primary marrow. As yet, no endochondral bone formation. (From Maximow and Bloom after Sobotta.)

for the conveyance of the blood vessels and excavate small cavities. The excavations in which the osteoclasts lie are known as *Haversian lacunae*. Osteoblasts which have advanced into the interior with the blood vessels cause the formation of true osseous tissue in the walls of the spaces formed by the osteoclasts.

At a somewhat later date than that at which the primary ossification center appears in the shaft, a secondary or epiphyseal center appears in one or both ends of the bone. The calcification process and subsequent ossification follow the same course as that described for the diaphysis. The epiphyseal and dia-

physeal areas, however, remain separated from one another by a layer of uncalcified cartilage, the *epiphyseal plate*, until a certain age, which varies between different bones.

Through the combined action of osteoclasts and osteoblasts a complete replacement of the calcified cartilage results and the structure characteristic of bone gradually evolves. The center of the shaft becomes hollowed out to form the medullary canal. The spaces which have been formed by the osteoclasts in the shaft of the bone itself become joined together and constructed into the system of *Haversian canals* which serve as conduits for the transmission of blood vessels. In the ends of the bones the spaces are much larger and irregular, and becoming filled with red marrow constitute the characteristic spongy or cancellous bone of this region. The walls of these spaces appear in cross-section as interlacing bars of osseous tissue and are usually referred to as *trabeculae*. The bone forming the walls of the Haversian canals is laid down in a series of concentric tubular lamellae. As each lamella is completed it imprisons the osteoblasts within small lacunae from which numerous fine canals are given off, into these the processes of the osteoblasts penetrate. The osteoblasts in these situations lose their osteogenic function but do not disappear. In the developed bone they are referred to simply as *osseous cells* or *osteocytes*.

A long bone grows in length at the junction of the epiphyses with the diaphysis and in thickness through the activity of the osteoblasts of the deeper layers of the periosteum. The Haversian canals and the marrow cavity are also lined with a membrane—the *endosteum*—containing osteoblasts through which increased width of the bone is also brought about.

The osteoblasts and osteoclasts are concerned not only with the development and growth of bone, but are active throughout life and are responsible, it is believed, to a large extent at any rate, for the lability of adult osseous tissue (p. 863). Healthy bone is constantly being broken down, resorbed and repaired. Several conditions may alter the balance one way or the other between these two processes, e.g., the relation of the calcium intake to the calcium requirement and the activities of various ductless glands—parathyroid, thyroid, pituitary, etc. Large numbers of osteoclasts are in evidence when bone resorption is taking place. Osteoclastic activity is therefore pronounced when remodelling of bone is occurring, as in the removal of excess callus or in the restoration to normal dimensions of the enlarged end of a bone in healing rickets, in certain bone diseases, and in wasting diseases. In old age also, the resorptive process outstrips the processes of repair, the bones become rarefied (*senile osteoporosis*) and more fragile. Ham and some other authorities deny that the osteoclasts are active agents in the removal of the bone tissue, claim-

ing that appearance of these cells is merely incidental or sequential to the resorption process

When bone formation is in the ascendancy, as during the repair of a fracture, the osteoblasts show active proliferation

Newly formed bone is stained selectively *in vivo* by madder, a red dye, or by alizarin, a derivative of it. These are therefore valuable agents for the investigation of the growth of bone in the living animal or of the action of parathyroid hormone and other influences upon bone metabolism. The dye, which is given orally, is deposited with the mineral constituents, calcium and phosphorus

A CONSIDERATION OF THE FACTORS UNDERLYING THE CALCIFICATION PROCESS Several theories have been advanced in attempts to picture the processes underlying the deposition of calcium salts in the cartilaginous matrix. The four principal theories are as follows

(1) A protein constituent of cartilage, it has been supposed, adsorbs calcium, for which it exhibits a special affinity, the calcium subsequently combines with phosphorus to form tricalcium phosphate. Wells showed, for example, that bone salts were deposited in a piece of boiled cartilage placed in the abdominal cavity

(2) The saturation of a solution with a salt such as calcium phosphate and the precipitation of the latter in solid form is determined by the product of the concentrations of the Ca^{++} and PO_4^{--} ions in the solution, i.e., upon the *ion product* $[\text{Ca}^{++}] \times [\text{PO}_4^{--}]$, and not upon the total quantities of calcium and phosphorus present. The ion product at which the solution is just saturated is called the *solubility product*. The presence of protein in the solution reduces the degree of ionization, part of the calcium, as we have seen, becoming bound to form calcium proteinate, a weak electrolyte. The body fluids are therefore capable of dissolving more calcium phosphate than is a protein-free solution similar in its salt composition to that of plasma. The CO_2 dissolved in the body fluids also increases their ability to hold calcium salts in solution. Any reduction in the protein concentration or in the CO_2 tension in the body fluids would therefore be expected to favor the deposition of calcium salts

An explanation of the calcification process upon a basis afforded by the foregoing considerations has been advanced by Howland. It is assumed that the fluids bathing the cartilage cells have, in common with other extravascular fluids, a lower protein concentration than has plasma. It is fur-

ther suggested that since the cartilaginous matrix has a low metabolism, the CO_2 tension of the fluids bathing its cells is lower than that of plasma, the pH of these fluids will therefore tend to be higher. Such conditions it is argued must favor the deposition of calcium salts. A low concentration of inorganic phosphorus and of calcium in the plasma will, on the other hand, tend to retard calcification, this would be completely arrested if the ion product were below the value at which precipitation occurs. In practice Howland and his associates have employed the product of the *total* calcium and inorganic phosphorus, each expressed in milligrams per cent, as an index of the calcification process in infantile rickets. They state that in children with active rickets the product is practically always below 40, and when the disease is severe, below 30, whereas in normal children it is between 50 and 60 (i.e., $\text{Ca } 11 \times \text{P } 5 = 55$). It is obvious that a fall in either inorganic phosphorus or calcium would tend to lower the $\text{Ca} \times \text{P}$ product, actually two varieties of rickets were distinguished—a low phosphorus and a low calcium type. It is now recognized, however, that these are simply different stages of the disease and that the level of the inorganic phosphorus alone is a more useful criterion by which to judge the extent of the calcification defect than the $\text{Ca} \times \text{P}$ product. For example, when rickets is progressing, the inorganic phosphorus of the plasma is lowered to between 3 and 4 mg. or less per cent, but the serum calcium is not far from the normal level. During the stage of healing, i.e., of active calcification, the inorganic phosphorus tends to rise and the serum calcium to fall (pp 851 and 853), there might therefore be little change in the $\text{Ca} \times \text{P}$ product

Another factor considered to be of importance in the calcification process is the *supersaturation* of the body fluids with calcium salts. That is, quite apart from the greater solubility of calcium salts in fluids containing protein and CO_2 , the concentrations of calcium phosphate and calcium carbonate in the body fluids are constantly maintained above their saturation limits, owing to the extreme slowness with which final equilibrium between the solid and liquid phases is established. Precipitation of these salts from solution will continue so long as the ion product is above that of the saturation level

Calcification, however, cannot be explained upon a physico-chemical basis alone, the activity

of living cells is also involved in the process. Shipley, Howland and Kramer showed, for example, that in experiments *in vitro* calcification was inhibited by a protoplasmic poison such as HCN. Two views have been expressed as to the nature of the vital processes concerned.

(3) Watt, from a comparison of the shapes of calcium phosphate particles precipitated in certain inert colloids with those formed in bone, concluded that calcification was not a simple precipitation of calcium phosphate from solution but was due to the active *secretion* by the osteoblasts of bone salts derived from the calcium and phosphorus of the blood.

(4) According to Robison and his colleagues, calcification is primarily dependent upon enzyme action through which the fluids in immediate relation to cartilage cells become highly supersaturated with calcium phosphate. These observers have accumulated much evidence in favor of their view. They have demonstrated the presence in bone (and also in plasma and other tissues, see below) of an enzyme capable of hydrolyzing various phosphonic esters, e.g., hexose-monophosphate, glycerophosphate, etc. This enzyme is called *phosphatase*. It is believed to be a product of the osteoblasts, the proliferating cartilage cells and the cells of the inner layer of the periosteum. According to this conception of the calcification process the enzyme liberates inorganic phosphate from phosphonic esters and raises, locally, the concentration of the PO_4^{3-} ion. The product of the concentrations of the Ca^{++} and PO_4^{3-} ions then exceeds the solubility product of calcium phosphate, which is in consequence deposited in the cartilaginous matrix.

Evidence bearing upon this hypothesis is as follows

(a) It was shown by Robison that when the head of a bone from a rachitic rat was immersed in a solution of calcium hexosemonophosphate at body temperature a deposit of calcium phosphate occurred in the zone of preparatory calcification (p. 764). This was attributed to the liberation by phosphatase of inorganic phosphate from a phosphonic ester, thus raising the product of the concentrations of PO_4^{3-} and Ca^{++} ions.

(b) Shipley, Kramer and Howland found that calcification of a rachitic bone occurred if placed in normal serum. Calcification also resulted if the bone were placed in a sterile solution of inorganic salts containing sodium chloride, sodium bicarbonate and magnesium sulphate, together with calcium and inorganic phosphate in the same concentrations as in normal serum. They concluded that living processes were concerned,

since calcification was inhibited by HCN. They believed, however, that phosphatase could have played no part in the process for the artificial solution did not, of course, contain a phosphonic ester.

(c) Robison claimed that the result of the experiment just described was dependent upon the fact that in the solution used, the concentration of the calcium phosphorus compound was near the point at which spontaneous precipitation might be expected to occur. In normal plasma, as already mentioned, calcium and phosphorus remain in solution at these concentrations because it contains protein which depresses the ionization of calcium, the ion product upon which precipitation depends is therefore considerably lower. Robison and Soames showed that calcium phosphate precipitates after a few days from a solution such as that employed by Shipley and associates if simply allowed to stand. It was found indeed by the latter workers that the addition of protein to the extent of 1 or 2 per cent, inhibited the calcification of the immersed bone. Robison and Soames showed later that though calcification of rachitic bone will occur when immersed in a solution containing calcium and phosphorus if the concentrations are sufficiently high, i.e., $\text{Ca} \times \text{P}$ product over 40 (4 mg P and 10 mg Ca per 100 cc.), calcification will not occur if the product is lower than this unless a phosphonic ester is added. Quite small amounts of the ester (glycerophosphate) were sufficient to cause calcification.

(d) Phosphatase is present in bone in largest amounts when and where active calcification is taking place (see below) which strongly suggests that it plays an essential rôle in the calcification process.

(e) Certain facts have been cited as opposed to the phosphatase hypothesis: (a) in rickets the phosphatase activity of bone and plasma is increased rather than the reverse, and in the blood of rachitic rats the percentage of phosphonic esters hydrolyzable by phosphatase is not below normal. (b) Normal plasma contains only very small quantities of phosphonic esters, which seem inadequate to supply the inorganic phosphate necessary for the calcification process. (c) Certain tissues which do not calcify normally are rich in phosphatase while other tissues such as the arteries, which are frequently the site of pathological calcification, do not contain the enzyme.

Robison does not contend, however, that the phosphatase mechanism is the only one concerned in the calcification process for, as mentioned above, calcification will occur in the absence of a phosphonic ester provided the concentrations of calcium and inorganic phosphate are sufficiently high. He and his colleagues found that when bone slices were treated with KCN or with certain fat solvents (alcohol, chloroform or acetone) before placing them in the supersaturated solution, calcification did not occur in the absence of glycerophosphate. These substances, however, exerted little or no inhibitory effect if glycerophosphate were

present, i.e., the phosphatase mechanism was not paralyzed and calcification proceeded. Formalin, on the contrary, prevented calcification whether a phosphoric ester was present or not. These results, in Robison's view, point to two distinct mechanisms governing the calcification process (a) The phosphatase mechanism, poisoned by formalin, which produces in the fluids bathing the cartilage cells a state of supersaturation in respect to bone salt. (b) A mechanism poisoned by several agents, especially cyanide, which is responsible for the deposition of bone salts from a supersaturated solution, whether this is the result of phosphatase action or is brought to the cartilage matrix from another source. The nature of the second mechanism is unknown. It may be due, Robison suggests, to a "slight increase in the pH of the matrix fluid brought about by some membrane equilibrium." Since this mechanism is inhibited or paralyzed by cyanide it is evidently dependent also upon the activity of living cells.

The distribution and properties of bone-phosphatase—“alkaline” phosphatase

Phosphatase is present in greatest amounts in ossifying cartilage, in smaller amounts in formed bone, but is absent from resting epiphyseal cartilage and from non-ossifying cartilage in other situations. It was shown by Robison to be absent from the patella before the appearance of the ossification center in this bone but present thereafter. The teeth of young animals contain it in relatively large amounts. It is present in milk⁵ and also in the floral parts of plants.

The optimum pH for phosphatase activity is around 9.0. Magnesium ions greatly increase the activity of the enzyme, whereas calcium ions are mildly inhibitory. Phosphatase activity has been demonstrated in a number of tissues (see table 83). The “alkaline” phosphatase in plasma, kidney and intestine are probably identical with that found in bone.⁶ Bone is apparently the main if not the sole source of plasma phosphatase, since this is not appreciably reduced after the removal of various organs (intestine, kidney, spleen, pancreas, etc.). Phosphatase is excreted by the liver, a marked rise in plasma phosphatase therefore occurs in obstructive jaundice and in jaundice

due to liver damage, but not in the purely hemolytic type (chapter 40).

The following phosphoric esters are hydrolyzed by bone phosphatase—hexosediphosphoric ester (of Harden and Young), phosphopeptone from casein, guanylic and adenylic acids of yeast

TABLE 83*

Relative phosphatase activity of tissue extracts prepared under similar conditions from various mammalian tissues

Tissue	REFERENCE					
	Forral (1923)	Robison (1923)	Kay (1928, 1)			Kay 1931 2
	Man	Young rabbit	Adult			Adult rat—average of 4
			Rabbit	Cat	Man	
Intestine	100					
Duodenum			50†	93†	57†	46
Jejunum			100†	100†	85†	33
Ileum			53†	81†	100†	15
Colon			17†	34†	27†	6
Kidney	58	36	33	38	35	100
Ossifying cartilage		100				
Whole bone			20	10		76
Liver	16	43	12	4	6	4
Pancreas	8	11				
Lung			10	26	7	20
Blood		14				1
Testis	12					13
Brain cerebrum			3	3	4	7
Cardiac muscle			1	1		5
Skeletal muscle		4	1	1		2
Artery				Nil	Nil	

* Modified from H. D. Kay, *Phys. Rev.* 1932, 12, p. 388.

† Mucosa only.

The figures in each column are relative to the other, but the different columns cannot be compared quantitatively.

nucleic acid, pyrophosphoric acid, and fructose phosphoric acid. Inosinic acid and pyrimidine nucleotides (ch. 48) are hydrolyzed by intestinal and kidney phosphatase and also probably by bone phosphatase.

The pathological calcification of arteries such as the aorta, which Kay states does not contain the enzyme, cannot be satisfactorily explained upon the phosphatase hypothesis.

The phosphatase in bone and the other solid

⁵ Kay and Graham have introduced a test by which one may determine whether a given sample of milk has been properly pasteurized. The test is based upon the fact that the temperature used in the pasteurization process destroys the activity of the enzyme.

⁶ Phosphatase activity is expressed as the number of mg. of inorganic phosphorus liberated per gram of tissue from sodium β -glycerophosphate after 48 hours hydrolysis at the optimum pH and at a temperature of 38°C.

tissues mentioned is reduced by the administration of irradiated ergosterol in amounts which cause the withdrawal of calcium from the bones and calcification of the tissues, whereas small doses cause an increase. On the other hand, it has been shown by Kay that the plasma phosphatase is increased, often markedly, in diseases involving extensive changes in bone structure (see table 84).

Certain tissues, e.g., kidney, serum and semen, but especially the human prostate, contain an "acid" phosphatase. Its optimum pH is between 4 and 5.4. Its concentration in the serum is increased in prostatic cancer with metastases.

DEFECTS OF OSSIFICATION AND PATHOLOGICAL CALCIFICATIONS

Diseases of bone

Several of these have been considered in other parts of the text—*infantile rickets* in chapter 55 and on pages 851 and 865, *late rickets* and *osteomalacia* on page 768, and *osteitis fibrosa cystica* on page 858.

The hardness, strength and rigidity of healthy bone depend upon the proportions of the organic and inorganic constituents incorporated into its structure, much as the properties of a plaster bandage depend upon the impregnation of the cotton mesh with plaster of Paris. The mineral and fibrous components are of equal importance, each reinforces the other. The cotton bandage has a certain tensile strength but lacks rigidity, a cast of plaster of Paris alone has maximum rigidity, but is brittle and readily broken or crushed. In most bone diseases the normal proportions between these two components are altered. In rickets and osteomalacia, for example, the bone salts are reduced in relation to the organic material. In these diseases, as also in *osteitis fibrosa cystica*, the bones are in consequence softer and more yielding than the normal. In certain other bone conditions the proportion of mineral to organic material is increased. The bone as a result is brittle and easily fractured. In other instances again there may be little change in the proportion of these two materials but the mass of the bone is increased or diminished with corresponding variations in strength.

Osteitis deformans (Paget's disease) is a disease of the skeleton involving mainly the bones of the skull, pelvis, limbs and spine. The cranium is enlarged and its wall greatly thickened, the long bones of the limb are massive and curved, the back is bowed (kyphosis) and its movements restricted. The organic matter of the bones is in-

creased and the calcium content decreased, but the total amount of phosphorus is not far from normal. A pronounced degree of arteriosclerosis is frequently a feature. There is good evidence that the overgrowth of osseous tissue is secondary to a primary bone destruction. Metabolic studies in this disease have yielded little information though there is said to be a retention of calcium and phosphorus.

TABLE 84*

Changes in the phosphatase content of the plasma in disease

DISEASE	NUM BER OF CASES	PHOSPHATASE CONTENT OF PLASMA		
		Highest	Low est	Mean
		unit	unit	unit
Arthritis without bony changes	11	0.33	0.11	0.17
Arthritis with bony changes	7	0.25	0.09	0.17
Exophthalmic goiter	7	0.75	0.27	0.47
	8†	0.53†	0.19†	0.36†
Osteomyelitis	8	0.41	0.14	0.27
Fragilitas osseum (in infants or children)	6	0.66	0.16	0.41
Acromegaly	2	0.32	0.22	0.27
Rickets (infantile)‡	13	1.7	0.42	1.03
Rickets (renal)	2	1.5	0.9	1.2
Adolescent rickets	1			>2.4
Osteitis fibrosa (generalized)	3	>2.5	1.5	>1.8
	3†	1.8†	1.06†	>1.31†
Osteitis deformans	24	3.4	0.65	>1.7

* Modified from H. D. Kay, *Phys. Rev.*, 1932, 12, 412.

† Hunter (1930).

‡ Average for normal infants of approximately same age = 0.32 arbitrary unit. Average for normal adults = 0.10–0.21 unit.

Fragilitas ossium (osteogenesis imperfecta) is a congenital disease characterized by thinness and extreme fragility of the skeleton, especially of the long bones and ribs. The cranium shows defective calcification. Fractures result from the most trivial injuries or may occur without any apparent cause. Union and healing of the fractures occur, however, as readily as in a normal bone. The bones have a low calcium content, the cortex is very thin and the medullary cavity dilated.

Achondroplasia is a congenital disease in which endochondral ossification (p. 863) of the limb bones, especially the humeri and femora, is defective. Periosteal ossification is active. The long bones are therefore much shorter and thicker than normal, strong and dense. A characteristic type of dwarf results—short arms and legs with a trunk and head of almost normal dimensions. The bones of the base of the skull fuse

prematurely and the development of certain facial bones is abnormal. Achondroplastic dwarfs develop as a result of these abnormalities a distinctive facies—depressed nasal bridge (pugnose), broad forehead and prominent lower jaw.

The cause or causes of the three foregoing osseous abnormalities is unknown. The possibility of some endocrine disorder, of course, comes to mind, but there is little or no evidence of such.

Marble or chalky bone (Albers-Schönbergs disease)
In this condition the density of the bone is greatly increased, the cancellous tissue is filled with a chalky material and the medullary canal may be almost obliterated by the concentric thickening of the shaft. The excessive calcification, however, renders the bones soft and brittle. Calcification of soft tissues—arteries, lungs, tendons—is often a feature. The disease is exceedingly rare and is mentioned here only because of the interesting possibility that some abnormality of parathyroid function is responsible. A case has been reported by French observers in which there was enlargement of the parathyroids, and Selye has produced in rats a state of increased bone density, which he considers comparable to this disease, by the administration of parathyroid extract after the animals had become tolerant to the usual action of the hormone.

Renal rickets, renal osteitis fibrosa cystica or generalisata
This is a condition commencing most usually in childhood and associated with chronic nephritis, rarefaction of the skeleton, dwarfism, low serum calcium and sometimes, calcium deposits in the soft tissues, especially in the kidneys, there are also acidosis and phosphate retention. The parathyroids show hyperplasia. We have seen that a rise in serum phosphate causes a reciprocal reduction in serum calcium and it is usually held that the hypocalcemia seen in this disease is due to the retention of phosphate resulting from the renal insufficiency. The hypocalcemia so produced causes, it is presumed, a drain of calcium from the bones. Increased excretion of phosphate into the intestine as a result of the diminished excretion by the kidney, with consequent depression of calcium absorption (p. 861) is possibly a contributory factor in the production of the low serum calcium. Albright and his colleagues suggest that renal osteitis fibrosa generalisata (p. 858) is a more appropriate name for this disease than renal rickets inasmuch as the histological changes in the bones are indistinguishable from *primary* hyperparathyroidism. This brings up the question as to whether the demineralization of the skeleton is due to the hyperparathyroidism induced as a compensatory reaction to the low serum

calcium. In the opinion of Albright and his associates, the withdrawal of calcium from the skeleton is not directly due to the parathyroid hyperplasia, i.e., to the parathyroid hormone itself, but to the acidosis which results from the failure of the diseased renal tubules to produce ammonia. Calcium is used as a base for the neutralization of acid, the tendency toward a fall in serum calcium is met by the withdrawal of calcium from the bones. Parathyroid hyperplasia occurs as a compensatory reaction to the hypocalcemia. The effect upon the bones was found to respond to measures which reduce the acidosis.

Calcification of soft tissues

DYSTROPHIC CALCIFICATION is the term applied to the deposition of calcium salts in dead, dying or chronically inflamed tissues and in areas of fatty or hyaline degeneration. Thus areas of necrosis, infarcts, scar tissue, caseous tuberculous areas and degenerated nerve cells, tend to undergo calcification. Calcification also occurs in the infarcts of the placenta which appear in the later half of pregnancy. Many of the examples of pathological calcifications to be described are simply special examples of dystrophic calcification. The factors determining the deposition of calcium salts in devitalized tissues are obscure. It has been suggested that since the CO_2 production in such tissue is minimal or entirely absent they will have a more alkaline reaction, this, of course, would tend to cause the deposition of calcium salts.

CALCINOSIS is the name given to conditions in which (a) calcified areas are scattered throughout the skin and subcutaneous tissues (*calcinosis circumscripta*), or (b) a more generalized calcification of skin, interstitial tissues, tendons, fascia or muscles occurs (*calcinosis universalis*). When the calcification process involves predominantly the interstitial tissues of the muscles the condition is usually referred to as *myositis ossificans*. In calcinosis the calcium and phosphorus levels of the blood are normal. Metabolic studies have in some instances revealed a retention of calcium. In the region of the calcified areas true bone formation may occur. Calcinosis of the superficial tissues is in many cases associated with scleroderma (a condition characterized by induration of the skin due to an increase in the intercellular collagenous tissue). The cause of calcinosis is obscure, the calcification process may be secondary to degenerative changes in the tissues themselves. There is no evidence that it is dependent upon an abnormality of parathyroid function, though Selye has reported a condition in rats resembling scleroderma following the administration of parathyroid extract.

ARTERIAL CALCIFICATION (1) *Arteriosclerosis* is seen in two main forms (a) the *atherosclerosis* (atherē = crushed grain, porridge) of *Marchand*, (b) the *medial sclerosis of Mönckeberg* (see ch 16)

METASTATIC CALCIFICATION This term connotes a transference of calcium from the skeleton to the soft tissues. It occurs in animals treated with excessive doses of parathyroid extract or irradiated ergosterol. Though the calcium deposits may be found in any of the soft tissues, the arteries, kidneys and lungs are especially susceptible to calcification. The fundus of the stomach is also a common site. It will be noted that the three last-mentioned organs eliminate acid, and it has been suggested that since this will leave the cells more alkaline in reaction, a condition favorable to calcium deposition is created. Metastatic calcification also occurs occasionally clinically. It has been reported in hyperparathyroidism (p 858), renal rickets and in certain bone diseases, e.g., multiple myelomata. It is very natural to assume that in conditions of disturbed calcium metabolism and destructive disease of bone the calcium deposits are simply the result of the excess calcium in the circulation. It is quite possible, however, that, in some instances at any rate, it is secondary to tissue injury and may therefore be, in reality, a type of dystrophic calcification induced by a toxic agent. Parathyroid extract and irradiated ergosterol, for example, besides their effects upon calcium metabolism have a definitely toxic action. Furthermore, metastatic calcification is in some instances associated with hypocalcemia.

It is a fact of great interest that the calcium deposits in the arteries and in other soft tissues in the various types of pathological calcification have the same composition as the main mineral compounds of bone. Evidence obtained by both chemical and physical methods support this conclusion. In certain instances actual bone formation occurs, even to the extent of producing red marrow tissue. Areas of ossification have been observed in the aorta and in the neighborhood of calcium deposits in the necrotic kidney of the rabbit. Also, as shown by Huggins, if a section of the mucosa of the bladder be transplanted into the subcutaneous tissues it becomes the site of bone formation.

Renal calculi (nephrolithiasis, urolithiasis)

Kidney and bladder stones composed largely of calcium phosphate have been produced in experi-

mental animals (rats) by the administration of irradiated ergosterol or parathyroid extract, in excess. They are apparently the result of the excretion in the urine of large amounts of calcium liberated from the bones.

Renal or vesical calculi are also associated with various bone diseases of a destructive nature, they are quite frequent in hyperparathyroidism. Renal calculi may be composed of calcium oxalate (in acid urine) or of calcium carbonate or phosphate (in alkaline urine) or of urates, uric acid or cystine. There is an undoubted relationship between the urinary excretion of citrate and the production of calcium stones. Citrate excretion is reduced in those suffering from renal calcium deposits even though the intake of citrate is greatly increased. In normal persons the urinary excretion of citrate runs parallel with the calcium excretion and varies with the calcium of the diet. It is suggested that citrate in some way reduces the tendency toward the precipitation of calcium. Nephrolithiasis is particularly prevalent in the tropics and the possibility has been suggested that hypervitaminosis D, due to over-irradiation with ultraviolet light and, possibly, deficiency of vitamin A are causative factors. The production of renal calculi in animals by hypervitaminosis D, as mentioned above, occurs only, however, when doses are employed which cause bone resorption and an increased concentration of calcium in the urine. Such results have therefore little bearing upon the question of the production of urinary calculi in the human subject. Also, though there is some evidence that avitaminosis A is conducive to the development of urinary calculi in animals (the cornification of the epithelium of the urinary tract being, apparently, a predisposing factor), there is little warrant for applying the results of animal experiments to the question of urinary lithiasis in man.

Albright's syndrome (polyostotic fibrous dysplasia, osteitis fibrosa disseminata). This is a rare disease with bizarre manifestations, consisting of bone cysts, fractures, patches of brown pigmentation in the skin, outward bowing of the femur and, in female children only, precocious puberty. The bony changes resemble those of osteitis fibrosa cystica but are more localized, they, as well as the cutaneous pigmentation, often show a segmental distribution, suggesting a neurological or developmental origin. There is no evidence of parathyroid overactivity or of a fault in any other endocrine

CHAPTER 61

THE ENDOCRINE ORGANS OF SEX THE THYMUS AND PINEAL GLANDS

THE CHARACTERS OF SEX

The sex glands—testes or ovaries—are known also as the *gonads*. They are the *primary organs of sex*. They furnish the male or female sex cells (spermatozoa or ova) and the hormones upon which the *ultimate* maleness or femaleness of the animal depends.

The sex, and so the type of sex gland, which is to develop is ordained, however, at the earliest possible time in the history of the individual, namely, when the parental ovum and spermatozoon conjugate. The determining factor is the type of chromosome (X or Y) in the sperm cell. The cells of the human body possess 48 chromosomes—24 pairs. In males the chromosomes of one pair are dissimilar. One is small and designated Y. The other chromosome of this pair is large like those in the other pairs and is referred to as X. In the female, the sex chromosomes are both large, that is, X chromosomes. When the male and female germ cells mature each receives only half the number of chromosomes. Approximately half of the sperm cells will therefore contain a Y, and the remaining half or so will contain an X chromosome. An ovum, of course, can contain only an X chromosome. Fertilization of an ovum by a sperm cell possessing a Y chromosome, results in a cell (zygote) containing a chromosome of each type. The body cells of the individual to which this cell ultimately gives rise will therefore contain each a pair of XY chromosomes, the offspring will be male. If the ovum is fertilized by a sperm cell containing an X chromosome the offspring will be female (XX). Certain diseases and defects are linked with the X chromosome of the male. Among such sex-linked diseases are hemophilia and color blindness. They appear in the male but are transmitted by the female who does not herself show the disease or abnormality. For example, a man afflicted with hemophilia if he marries a normal woman does not transmit the disease to his sons (XY), for the X chromosome received from the mother does not carry the defect, nor does the Y chromosome derived from the father. The daughters, however, have received an abnormal character in the X chromosome

from the father as well as a normal one from the mother. But, since the abnormal character is recessive, it is "suppressed" by the dominant normal character, and the daughters are free from the disease, though their cells must contain the X chromosome carrying the defect. When therefore the daughters marry and an ovum is fertilized by a sperm cell containing an X chromosome the daughters of the next generation again show no abnormality, but should a Y sperm fertilize the ovum and an XY (male) zygote result, the child will show the disease if one of the abnormal X chromosomes of the mother has paired with the Y from the father. The Y chromosome does not offset the effect of the abnormal X chromosome. It should be remembered that since the mother possesses two X chromosomes, a normal and an abnormal one, it is an even chance which one the child will receive.¹

Those organs, other than the sex glands, which are essential for procreation, such as the external genitalia, as well as the uterus, Fallopian tubes and vagina of the female and the seminal vesicles and prostate of the male, are spoken of as the *accessory organs of sex*. Other characters of sex only make their appearance at the time of sexual maturity (puberty), e g, the growth of hair upon the pubis of the human male or female, the development of the mammary glands in women, the development of the antlers of stags, the distinctive plumage of birds and the psychic manifestations of sex in man and in animals. These are spoken of as the *secondary sex characters*.

EFFECTS OF EXCISION OF THE SEX GLANDS OR GONADS—CASTRATION—SPAYING

Removal of the gonads (castration) from a young animal prevents the mature development of the accessory sex organs and the secondary sex char-

¹ In man the ratio of live male births to female is 105:100. If abortions and stillbirths are taken into account, there is an even greater disparity at conception between male (XY) and female (XX) zygotes. The viability of males is considerably less than that of females, i e, there is a greater mortality of male babies. The lower viability of males persists in later childhood, so that more males are eliminated and females come to predominate in the adult population.

acters fail to make their appearance. The effects of castration upon the secondary characters are evident even in such simple animal forms as the earthworm and the hermit crab. Castration of male frogs prevents the appearance of the sexual changes which normally occur during the mating season. The thumb pad and the fore limb muscles do not hypertrophy, the clasping reflex cannot be elicited and the animal does not emit its characteristic croaking sound. Transplantation of the excised testicular tissue into another part of the animal's body prevents the occurrence of these castration effects. The sexual development of birds and mammals is profoundly affected by castration. The castrated Leghorn cockerel (capon), for example, has a greater proportion of body fat than the normal bird, while the comb, wattles and barbles, and the sex instincts do not develop. Development of the spurs and plumage, however, is not prevented. Corresponding effects of gonadectomy are seen in the young turkey-cock and in the young of other avian species. Removal of the ovaries from the young hen (spaying)² causes the development of spurs, a comb resembling that of the cock, and male plumage. The spayed duck assumes the plumage of the drake.³

The effects of castration upon young cattle, horses and stags are well known. Castration of young bulls causes an increase in size of the skeleton and a greater deposition of fat. The mature development of the accessory organs is prevented. The antlers of young stags do not develop after castration and female deer after atrophy or disease of the ovaries may develop horns. Castration of boys before puberty retards ossification of the epiphyses of the long bones with consequent enlargement of the stature due to disproportionate lengthening of the lower limbs. There is also adiposity, the fat tending to become feminine in distribution. The larynx is not prominent as in the mature male and the voice remains high-pitched. The hair fails to grow upon the face and body, but is unusually plentiful on the head. The penis remains infantile and sexual feeling is suppressed. Such a state is called *eunuchism*. Ovari-

² The terms ovariectomy and castration are also applied to removal of the ovaries. The latter term may be applied to excision of either testes or ovaries.

³ Male plumage is considered to be the basic or neutral type. Its development is suppressed by the female sex hormone. The female type of plumage then appears. Hence it is that male plumage appears after ovariectomy and persists after removal of the testes.

ectomy is followed by corresponding effects, if performed before puberty the characteristic feminine attributes do not appear, the girl tends to become somewhat mannish in type, the accessory organs fail to develop fully and menstruation does not occur. In animals subjected to this operation before puberty the estrus cycles do not ensue. If the operation is performed after puberty the estrous cycles are suppressed and the accessory organs atrophy. In women ovariectomy is followed by changes characteristic of the menopause (p. 893), namely, amenorrhea, atrophy of the sex organs and obesity.

Sex desire in higher animals and in the human subject is not, apparently, dependent entirely upon the gonads, for it is sometimes retained in

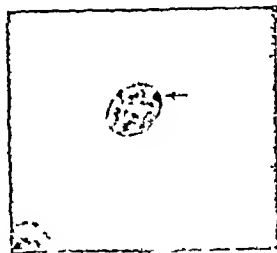


FIG. 61.1. Arrow points to sex chromatin in cartilage cell of cyphoid process of female (human). (After Moore and Barr.)

eunuchs and in women who have been ovariectomized after marriage—though removal of the ovaries before puberty, i.e., before the accessory organs of sex have reached maturity, or from virgin women after puberty, suppresses the sex libido. Stone found that a certain proportion of a series of male rats castrated after sexual maturity had been attained, pursued the female and were capable of copulating during a subsequent period of eight months. Rowe also cites the instance of a man in whom the sex libido was not abolished twenty-five years after the testes had been excised. Androgens derived from the adrenal cortex come to mind in this connection. Prepubertal removal of the testes prevents the development of sexual desire.

The sex chromatin. It was discovered by Barr and Bertram that the body cells of males and females could be distinguished from one another by microscopical examination. The nucleus of a female cell contains a minute dot of chromatin material, now known as the *sex chromatin*, which in the general tissue cells lies most usually close

to or in contact with the nuclear membrane. In nerve cells it is placed adjacent to the nucleolus, and in this situation has been termed the "nucleolar satellite" (see fig. 611, p. 872). Male cells do not, except rarely, possess a sex chromatin. Though proof is lacking it is believed that the sex chromatin is derived from the heterochromatic portion of the sex chromosomes (XX).⁴ The sex chromatin is formed in the earliest stages of development, it is of genetic origin and is uninfluenced by the sex hormones. The discovery of the sex chromatin has been put to an important practical use. Examination of the cells of the Malpighian layer of an excised scrap of skin from cases of hermaphroditism⁵ enables the dominant sex, when in doubt, to be determined, that is, to diagnose the type of sex chromosome (XY or XX) in the hermaphrodite's sex cells.

The free martin

The effect of the secretion of the gonads upon the sexual development of the embryo is shown in the case of twin calves when these are of opposite sexes. The male twin is always normal but the female, in the great majority of cases, shows abnormalities of the sexual organs. The uterus is small and underdeveloped, the clitoris is enlarged and penis like, the gonads are rudimentary and may resemble testicles and in some instances structures resembling vasa deferentia and seminal vesicles are present. The abnormal twin showing such intersexual characters is called a *free martin*. Keller and Tandler in Austria and Lillie in America have made a study of a number of such cases. They found that as a result of the fusion of the two chorions, a direct communication between the circulations of the male and female embryos exists. According to Lillie, the fusion of the chorions is seen at a very early stage of development, when the embryos are no more than 10 or 20 mm. long. In the rare instance in which the female twin was normal a communication between the two circulations did not exist. The masculinization of

the female twin is therefore attributed to the action of the internal secretion liberated by the gonads of the male. In order to explain the fact that masculinization of the female occurs but never feminization of the male, it is supposed that the hormone of the testis is elaborated at an earlier stage of development than is that of the ovary. Masculinization of female guinea-pigs has been induced by the injection of male hormone into the fetuses *in utero* (Dantchakoff), and Ivy and his colleagues produced corresponding changes in rat fetuses by means of male hormone treatment of the mother during gestation, feminization of male rats was also effected by antenatal treatment with female hormone (estradiol-dipropionate, p. 878).

Transplantation or grafting experiments

If the excised sex gland (testis or ovary) is transplanted to another situation in the body and survives, the otherwise inevitable effects of castration are prevented. This fact, first demonstrated by Berthold in 1849, proves conclusively that the sex gland (male or female) furnishes an internal secretion. Transplantation of the gland from its normal position into another situation in the same animal is called *autotransplantation*. The grafted tissue usually lives. This operation has been performed in the human female. The grafting of the gonad into another animal of the same species is called *homotransplantation*. This operation is not so successful, though a certain proportion of homotransplants survive and perform their functions, for a time at any rate. In this way feminization of the capon or masculinization of the bilaterally ovariectomized hen can be induced by the transplantation of the gonad of the opposite sex. Transplantation of the gonad into the body of a member of another species is known as *heterotransplantation*. It is the least successful type of transplantation. In some cases, however, histological proof of the survival of the transplant for some months has been obtained, but atrophy occurs eventually.

Development and structure of the ovaries

The ovaries as well as the testes arise from the coelomic epithelium covering the inner aspect of the Wolffian body. The epithelial cells of this region assume a columnar form, they proliferate to form several layers which constitute the *germinal epithelium*. The mesoblast underlying the germinal epithelium becomes thickened to form the *genital ridge*. Fingers of mesoblast grow upward into the overlying germinal epithelium while columns composed of the latter cells grow into the mesoblastic tissue. There is thus an interlocking of epithelium and mesoblastic cell masses. The epithelial columns, or *genital cords*, as they are called, later become broken up into *cell nests*. From these, the *Graafian follicles* of the ovary or the *convoluted seminiferous tubules* of the testes are developed. A layer of the germinal epithelium also comes to cover the surface of

⁴ The sex chromatin contains desoxyribonucleic acid, and is Feulgen positive, whereas the nucleolus contains ribonucleic acid and is Feulgen negative (ch. 48).

⁵ Hermaphroditism is divided into two main classes—*true hermaphroditism* and *pseudohermaphroditism*. In the former the individual possesses both testicular and ovarian tissue, either as separate gonads or combined as ovotestes. The pseudohermaphrodite has only one type of gonad—testis or ovary, but the sex abnormality is seen in the accessory organs of sex or in the secondary sex characteristics, or in both. Many cases of pseudohermaphroditism are of endocrine origin, the hormonal influence being exerted either in fetal life or postnatally (see adrenogenital syndrome, ch. 59). Sometimes certain regions, even one half of the body, are female while other regions or body half is male, such intersex individuals are called *gynandromorphs*.

the ovary The mesoblastic tissue surrounding the cell islands of the primitive gonad gives rise to the stroma and vascular tissue of the ovary or testis

Structure of the ovary The adult ovary is about the size and shape of a shelled almond and consists of a stroma of connective tissue which carries the blood vessels and in which are embedded a number of follicles—the *Graafian follicles*—in different stages of development. The stroma also contains a few smooth muscle fibers The surface of the ovary is covered with a layer of epithelial cells—the *germinal epithelium*—continuous with the epithelium of the general peritoneum. Irregular groups of epithelial like cells are also seen throughout the ovarian stroma of many animals Groups of these so-called *interstitial cells* are absent from the human ovary, though scattered cells similar in appearance, and probably also in function, are present

The Graafian follicles The germinal epithelium covering the ovary retains its embryonic character and proclivities after birth—until the end of the reproductive period (to about the 45th year in women) Its cells show active mitoses, as in the embryo, and multiply rapidly Columns of these cells penetrate the underlying ovarian stroma and become broken up into groups or islands in the ovarian connective tissue. One cell of the group develops into a primitive ovum around which the remainder become arranged in a circular row—the *granulosa cells* This body—the *primordial follicle*—which so far has no cavity and is about 0.5 mm. in diameter, migrates deeper into the stroma

The orderly arrangement of the cells in the development of the primordial follicles appears to be governed by the ovum itself, for when the ova are destroyed by X ray treatment, the germinal epithelium forms cords and clumps instead of follicles.

In the sexually immature animal, i.e., before puberty, only primordial follicles are formed Some 400,000 of these, according to one estimate, are present in the infantile ovary At puberty, as a result of pituitary activity, two more layers of cells are produced from the ovarian stroma. The inner layer—the *theca interna*—is thought to be formed by some influence exerted upon the stroma cells by the granulosa cells This layer is vascular and more cellular than the outer layer—the *theca externa*—which is fibrous in character At the same time that these changes are occurring, the granulosa cells multiply to form a cell mass several layers deep, now known as the *membrana granulosa* in which two concentric zones can soon be distinguished Later, these become separated by the collection of fluid (liquor folliculi) and a cavity, or *antrum*, is formed The cells of the inner zone, in immediate contact with and surrounding the ovum, show further multiplication and become heaped up into a mass known as the *cumulus oöphorus* or *discus proligerus* (fig 61.2)

As the follicle enlarges and the changes just described, which are spoken of as *follicular maturation* or *ripening*,

progress, it migrates again outwards, and when quite matured projects from the surface of the ovary Rupture of the follicular wall follows and the ovum is discharged By this time the ovum has enlarged to a diameter of about 70 micra, developed a deeply staining clear cell membrane—the *zona pellucida*—and undergone partial maturation (extrusion of the first polar body) After its discharge the ovum is conveyed along the Fallopian tube to the uterus The subsequent history of the Graafian follicle is given on page 880 Though

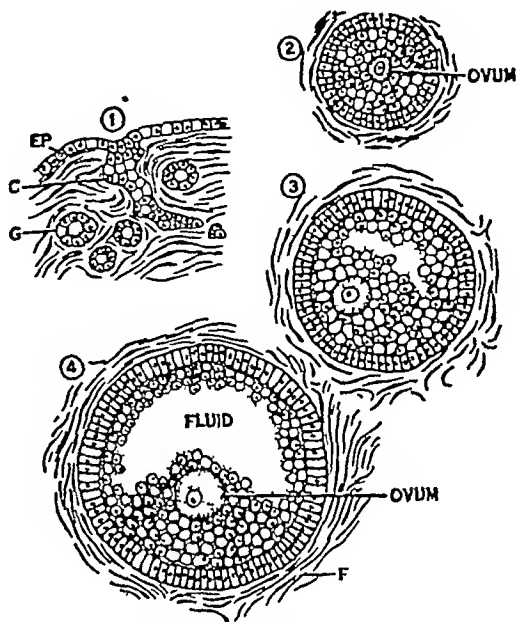


FIG 61.2 Successive stages in the development of the ovum and Graafian follicle (diagrammatic) EP, germinal epithelium covering the surface of the ovary, C, a cord of cells growing from the germinal epithelium into the ovarian substance, G, primordial ovum encircled by a single layer of (granulosa) cells In 4, the *membrana granulosa* and the *cumulus oöphorus* are shown but the follicle has not yet reached full development, F, ovarian stroma

large numbers of small follicles containing immature ova are present in the ovary from birth, ripening of the follicle and discharge of the ova do not occur until puberty In the mature ovary there appear at regularly recurring periods crops of primordial follicles But only a small proportion of these advance to complete maturation and discharge their ova. In women, no more than one or two ova are as a rule discharged each month or a total of 400 or so during the entire reproductive period The remaining follicles reach various stages in the ripening process and then undergo degenerative changes (*atretic follicles*) Each of the latter is finally replaced by fibrous tissue derived from the theca interna, a small scar (*corpus fibrosum*) alone remaining The factors which

determine the rupture of the ripe follicle are not definitely known. The accumulation of fluid and the consequent increase in intra-follicular pressure is probably a factor in some animals, or the contraction of the smooth muscle fibers of the stroma may play a part. In most species, including man, the follicle ruptures spontaneously, but in the rabbit, cat and ferret ovulation occurs only after copulation.

The structure of the testes and the development of the spermatozoa are described on p. 902.

THE SEXUAL LIFE OF FEMALES

Three sexual periods of different lengths occur in female mammals. (1) A single long period occupying the greater part of the animal's life

species. In the dog, for instance, two such periods of about six weeks' duration occur each year (spring and autumn), while in certain other species they occur more frequently, and in others again only once, and may be of long or of short duration. In the human, reproduction is not confined to any one part of the year, though a study of birth statistics shows that fertility is greatest at certain periods (April to June). It has been suggested that in primitive man a mating season corresponding to this time of year did exist.

(3) *The estrus cycles* These are periods of sexual activity in animals which occur once or oftener in each breeding season. The first cycle commences at puberty. In women, higher mon

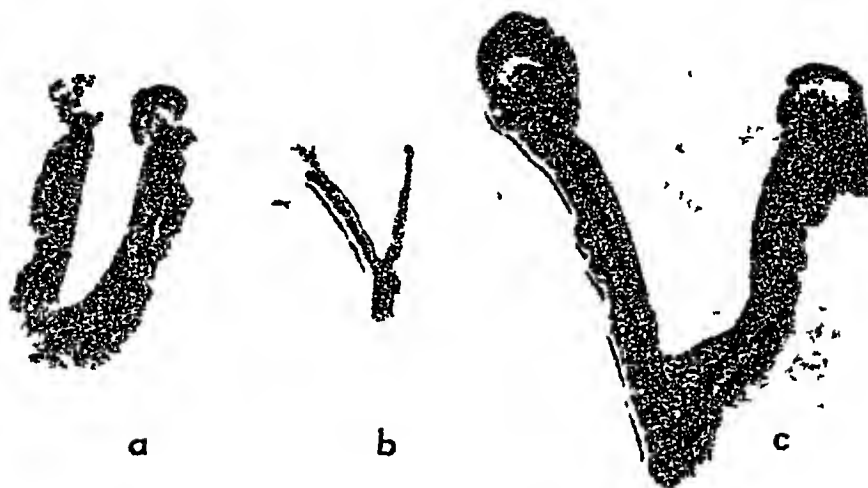


FIG 61.3 Showing uterus (a) in anestrus, (b), one month after ovariectomy, (c), in estrus (After d'Armour and Blood)

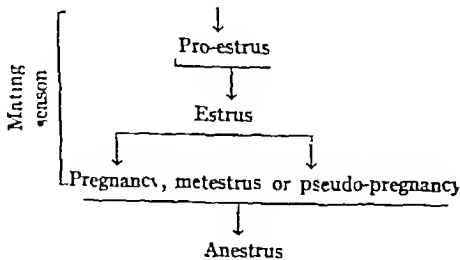
It commences at *puberty*, at which time the first ovulation occurs, the accessory organs of reproduction mature, sex desire is aroused and the secondary sex characters appear. It ends with atrophic changes in the ovary and accessory organs. In women the termination of this period is called the *menopause* (p. 893).

(2) A period which recurs once or oftener each year, known as the *mating, sexual, or breeding season*. In most species it is only during this time, or during a part of it, that the female will receive the male.⁶ The duration of the mating season and the number of times it occurs annually vary in different

keys and anthropoid apes they are represented by the menstrual periods. Animals such as the bitch, in which a single estrus cycle extends throughout the breeding season, are called *monestrus*. *Polyestrus* animals, on the other hand, are those such as the domestic cat, the mare, cow, sow, rat and mouse, in which two or more cycles, separated by short periods of sexual quiescence, occur in succession during the breeding season. In a monestrus animal the following phases of the estrus cycle are distinguished: (a) *Pro-estrus*, or period of "coming on heat". There are usually swelling and congestion of the external genitalia together with growth and increased vascularity of the uterus (fig. 61.3). There is, as a rule, some enlargement of the mammary glands and, in the dog and cow, bleeding from the vagina. During this stage of the cycle the Graafian

⁶ The corresponding period which occurs in the males of some species is known as the *rutting season*. In many species however, the male is capable of copulation at any time.

follicles are undergoing maturation preparatory to rupture (b) *Estrus*,⁷ or "period of desire" The female receives the male and ovulation occurs The term "heat" is commonly applied to the combined periods, pro-estrus and estrus (c) *Pseudo-pregnancy* or *pregnancy* The changes in the uterus initiated during the previous periods progress and in some animals, e.g., the bitch, rabbit and ferret, there occur pronounced proliferation and secretory activity of the uterine glands, hypertrophy of the mucosa and a great increase of the uterine blood supply It will be seen presently that these phenomena are dependent upon the formation of a corpus luteum. The growth of the mammary glands is stimulated This phase of the sexual cycle is known as pseudo-pregnancy The uterine changes, which are similar to those occurring in the pre-menstrual period of women are looked upon as anticipating the arrival of a fertilized ovum If this fails to ensue the newly-formed uterine fabric breaks down, the debris is discharged and the uterus returns to its resting state If impregnation of the ovum occurs the uterine changes persist and merge into those characteristic of the pregnant state. (d) *Anestrus* is the period of sexual quiescence between the mating seasons In the cycle of those animals in which pseudo pregnancy does not occur, the short period following estrus and during which the phenomena of estrus subside, is called *metestrus* The monestrus cycle may be illustrated as follows



In polyestrous animals the short intervals of quiescence separating the estrus cycles are called *di-estrus periods*. The term *anestrus*, as in the case of monestrous species, refers to the longer periods of rest between the mating seasons. Ovulation and, therefore, pseudo-pregnancy do not occur spontaneously in certain animals, e.g., cat, rabbit and ferret, but only after copulation. Though the onset of estrus in some species is quite

⁷ *Estrus*, L. = *gād fī*, with figurative meaning of frenzy or intense desire.

obvious, in rodents it is difficult to detect by the ordinary means. Stockard and Papanicolaou discovered that the vaginal mucosa of the guinea pig underwent certain changes (cornification of epithelium and mucification) coinciding with estrus. Similar changes occur in the vagina of the rat and

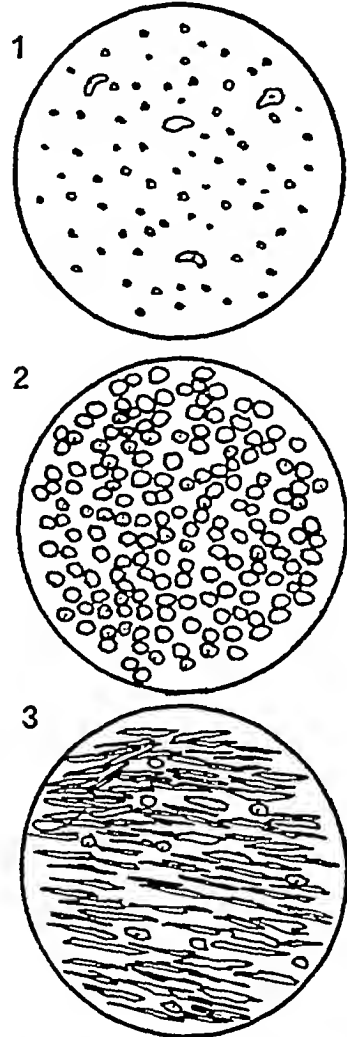


FIG 614 Vaginal smear, 1, in anestrus, 2, in early estrus, 3, in late estrus (After d'Amour and Blood)

mouse as well as of man. By the examination of a smear of the vaginal secretions the stages of the estrus cycle can be readily followed in these animals. Leukocytes which are present in smears taken during the diestrus or anestrus period are absent from the vaginal secretions during pro-estrus and estrus but large squamous (cornified)

cells appear (fig 61 4), with small pyknotic nuclei. The intact vaginal epithelium becomes stratified into a number of layers, with many of its cells filled with mucin.

OVARIAN HORMONES

These are (1) the *estrogenic hormone*, also known as the *follicular hormone*, or the *female sex hormone*, or chemically as *estradiol* (2) the *corpus luteum hormone*, *progestin*, or chemically as *progesterone*. The estrogenic hormone is concerned particularly with the first part of the estrus cycle, the corpus luteum hormone with the latter part (pseudo-pregnancy), and with the pregnant state (3) Evidence has been secured for the existence of a third ovarian hormone. This, which is known as *relaxin*, has been isolated from the corpus luteum.

THE ESTROGENIC HORMONE

The evidence derived from ovariectomy and transplantation experiments had made it clear that the ovary furnished an internal secretion which was responsible for the sexual development of the female, yet the results of experiments with ovarian extracts had been inconclusive until the work of Allen and Doisy in 1923. They obtained a potent ether extract from the liquor folliculi aspirated from hog's ovaries which was capable of inducing estrus in immature animals. The success of these workers was to a large extent due to the employment of a precise method for the demonstration of estrus, namely, the vaginal smear technique described above. Other investigators, for the most part, had relied upon uterine effects as a means of testing the activity of their extracts. The name *oestrin* (spelled *estrin* in the United States of America) was suggested by Parkes and Bellerby for this ovarian hormone.

The origin and distribution of estrin (oestrin)

The elements of the ovary responsible for the production of the follicular hormone are not definitely known. Its presence in high concentration in liquor folliculi and the fact that follicular maturation coincides with the onset of estrus, point to the follicular cells (probably of the theca interna) as being the chief source. Yet these cells cannot be solely responsible, for Parkes showed that estrus continues at regular intervals after the follicles have been completely destroyed by X-rays; moreover, the hormone can be extracted from the ovarian stroma alone. The interstitial cells in the

latter situation may be one source of the hormone. One of the most remarkable features of this hormone is its very wide distribution in animal tissues. It or a similarly acting estrogen is found in the blood, muscles, and urine of both pregnant and non-pregnant females, in the urine of adult males (estrone), and in the testes, the testes and urine of stallions being among the richest known sources (Zondek). It is present in very high concentration in the urine of pregnant women after the first 2 or 3 months. In the later months around 300,000 or more international units are excreted daily. It is also obtainable in large amounts from the urines of pregnant mares and monkeys. The human placenta also contains large quantities of estrogen which are believed to be actually manufactured by this organ, for women ovariectomized during the later months of pregnancy continue to excrete large amounts of the hormone in the urine. An estrogen is present in the corpus luteum, in the fetal membranes and in amniotic fluid, and in the adrenal cortex, it has been demonstrated in a human chorionic vesicle containing an embryo 12.5 mm in length.

Chemistry and terminology

Doisy and his colleagues and Butenandt, independently, isolated an estrogen in crystalline form from urine. Upon analysis it was found to have the empirical formula $C_{18}H_{22}O_2$ and to possess a ketone and a hydroxyl group. A second estrogenic compound containing one more molecule of water and having the formula $C_{18}H_{24}O_3$ was isolated from urine shortly afterward by Marran. This contains 3 hydroxyl groups but no ketone. It is less active than the previous form. Doisy gave the name *theelin* to the first of these estrogens, others speak of it as *ketohydroxyestrin* or *estrone*⁸ (*oestrone*). The second form, called *theelinol* by Doisy, is also referred to as *trihydroxyestrin* or *estriol* (*oestriol*), it is transformed to estrone by dehydration *in vacuo* with potassium bisulphate.

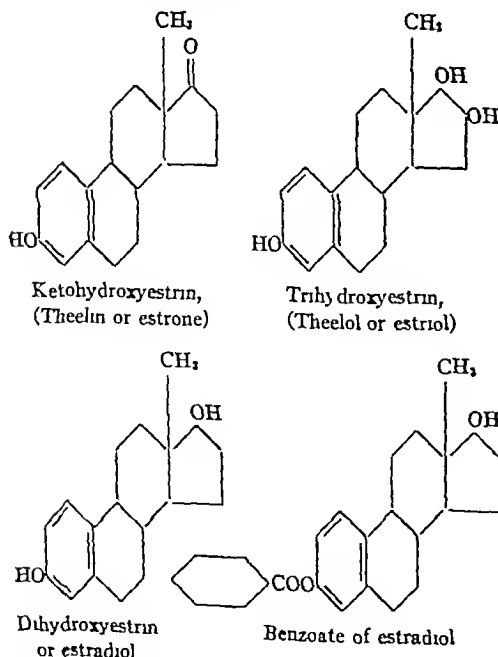
Estradiol (dihydroxyestrin, dihydrotheelin) ($C_{18}H_{24}O_2$) is a third crystalline compound which was first prepared in the laboratory by reduction of the ketone group in estrone to a hydroxyl group. There are three isomeric forms α , β and γ .

⁸ The advisory committee on nomenclature of endocrine principles appointed by the Council of Pharmacy and Chemistry of the American Medical Association recommend the following chemical names, 3-hydroxy-17-keto Δ^1 , Δ^3 , Δ^5 —estratriene, 3, 16, 17-trihydroxy- Δ^1 , Δ^3 , Δ^5 —estratriene and 3, 17-dihydroxy Δ^1 , Δ^3 , Δ^5 —estratriene, respectively, for estrone, estriol and estradiol. The

The alpha-form possesses much the greater activity of the three. Since its artificial preparation α -estradiol has been isolated by MacCorquodale and his colleagues from sows' ovaries and by Doisy and his associates from the urine of pregnant women, it is now generally believed to be the true estrogenic hormone, i.e., the estrogen naturally formed in, and secreted by the ovary. It is converted in the body to estrone and estriol. Winterstein and his associates have isolated the beta-form from the urine of pregnant mares. Iso-estradiol is a synthetic compound and has not been obtained from a natural source. The replacement of the hydroxyl group in the 3 position in estradiol with benzoic acid gives a product possessing a more prolonged physiological action than the original compound.⁹ This substance—the benzoate of estradiol—is known commercially by various names: *Estradiol-dipropionate* and *elinyl estradiol*, have even greater and more prolonged actions.

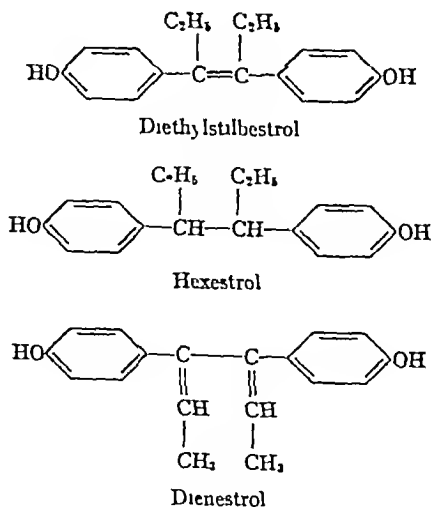
Emmenin, isolated by Collip and his associates from human placenta is an ester (probably the glucuronide) of estriol. It is active by oral administration.

The structural formulae of four of the estrogenic compounds just described are as follows:



The presence in the blood during pregnancy of large amounts of estrogen has always been something of a physiological puzzle, but some enlightenment has come through the discovery of Cohen and Marrian that estriol is excreted in the urine of pregnancy conjugated with glycuronic acid. This compound—the *glycuronide of estriol*—is physiologically inert, but is hydrolyzed, active estriol being freed, by the usual methods (involving acidification and heat) employed in isolating the hormone from urine. It was found, moreover, that the quantity of combined estriol in the urine of women just before the onset of labor became reduced, whereas that in the active (free) state became greatly increased. The significance of these findings in respect to the activation of the birth mechanism is obvious (p. 796). The conjugation of estriol with glycuronic acid at once suggests a detoxicating process in the liver whereby the uterus is protected from the action of an excess of circulating estrogenic hormone.

The hormones of the ovaries and testes belong to the class of sterols, being related chemically to the bile acids, to cholesterol and to calciferol. Starting with ergosterol, the synthesis of estrone was reported in 1936 by Marker and his associates. Others have synthesized estradiol from cholesterol. Estrogenic substances have been obtained from a wide variety of sources other than animal tissues and fluids. Among some of these sources are pe-



numbering of the "parent nucleus" estrane ($C_{18}H_{30}$) is the same as that shown on p. 838.

⁹ This is probably due to the fact that the benzoate as compared with the other compounds is less readily attacked and inactivated by the liver.

troleum, peat and lignite, yeast, rape seeds and pussy willows. Substances—benzanthracene compounds—with estrogenic properties and structural formulae suggestive of those given above have also been synthesized. The synthetic estrogenic

compound, 4 4'-dihydroxy- α - β -diethylstilbene, known generally as diethylstilbestrol, was introduced by Dodds and his associates in 1938 and has since come into clinical use. This substance exhibits an estrogenic potency between two and three times greater than that of estrone, its activity is reduced by only 60 per cent when given orally.

Two closely similar synthetic compounds also suitable for clinical use are *dihydroethylstilbestrol* or *hexestrol*, and *dieneestrol*. The latter has about 60 per cent of the activity of diethylstilbestrol, the potency of hexestrol is somewhat less again.

The relation between cancer and estrogenic substances

Some of the estrogenic benzanthracene compounds are capable of causing cancer when painted on the skin, and coal tar, which it will be recalled induces skin cancer, is also estrogenic. This relationship between carcinogenic and estrogenic action is highly suggestive. The production within the body (as a result of an abnormality in sterol metabolism) of some material of the benzanthracene type which may play a rôle in the genesis of cancer, is a possibility which comes to mind. "The cell proliferation which characterizes the estrous state is in some respects reminiscent of the early stages of malignant growth" (Cook and Dodds), so also is the development of the decidual tissue of early pregnancy which, as we shall see, is dependent upon the hormone of the corpus luteum. The following suggestive observations may also be cited in this connection.

(a) Corn found that ovariectomy of mice belonging to a strain showing a very high susceptibility to mammary cancer caused a marked reduction in the incidence of the disease.

(b) Lacassagne produced mammary cancer in male mice, which ordinarily are not susceptible to the disease, by the injection of large doses of an estrogenic hormone. This observer also states that differences exist between the estrus cycles of mice showing a high susceptibility to cancer and of those showing a low incidence of the disease. Loeb and his associates were unable, however, to detect any characteristic feature of the cycles in the various strains of mice which could be definitely related to the occurrence of cancer.

(c) In women cystic ovaries, and apparently hyperovarian function, have been found in frequent association with an hypertrophied endometrium.

(d) In monkeys prolonged treatment with estrone has resulted in atypical growths of the epithelium of the uterine cervix.

Method of assay and standardization of estrogenic activity

It has been the practice to assay the activity of estrogenic preparations upon ovariectomized rats or mice, a unit being defined as the minimal quantity of the material required to induce estrus, as determined by the vaginal smear technique, in the test animal. A rat or a mouse unit has not, however, an absolute and definite value but shows wide discrepancies between different laboratories, owing to variability in the details of the assay technique employed by individual workers. The Commission on Biological Standardization of the League of Nations has therefore defined a unit (international unit) of estrogenic activity as the activity of 0.1 microgram (0.1 γ) of crystalline estrone. The dosage of estrone required to produce a physiological effect is related to the body weight. In order therefore to induce a response in a woman possessing no endogenous supply of the hormone something like 500,000 international units would be required.

Actions of the estrogenic hormone

The main effects of this female hormone are as follows. (1) It induces estrus in immature animals and in ovariectomized adult animals, or in normal adult animals during anestrus, there result in consequence, hypertrophy of the uterus and proliferation of its glands (fig. 61.5), vaginal changes, growth of the mammary glands (p. 896), rhythmical contraction of the muscle of the uterus and Fallopian tubes, psychic and other phenomena associated with the estrus period. Prolonged treatment with estrogen causes pronounced hyperplasia of the endometrium which may come to resemble the clinical condition known as "Swiss cheese" endometrium. (2) It prevents the otherwise inevitable atrophy of the accessory reproductive organs in ovariectomized animals. (3) It is believed to be responsible for the development of the secondary sex characters which in some species are such prominent features of the mature female. In the male, estrogen in physiological dosage, or derived endogenously (testicular estrogen) appears to act synergically with androgen in the development of the secondary sex characters. Protracted treatment with estrogen causes hypertrophy of the fibromuscular tissues of the prostate, but has little or no effect upon the glandular elements. Growth of the male mammary glands is stimulated. In young males other feminizing effects are produced. (4) It prevents nidation of the fertilized ovum,

or induces abortion in early pregnancy (5) It causes hypertrophy of the adrenal cortex, as a result apparently, of stimulating the output of ACTH (6) This ovarian hormone is believed to be largely responsible for the enormous growth of the uterus during pregnancy When pregnancy is confined to one horn of the rabbit's uterus, that horn alone increases in size It is thought therefore that estrogen, which is in high concentration in the placenta, may act locally upon the uterine tissue rather than through the general circulation The estrogenic hormone does not stimulate the ovaries, continued injections actually reduce the size of the latter and of the testes, an effect attributed to

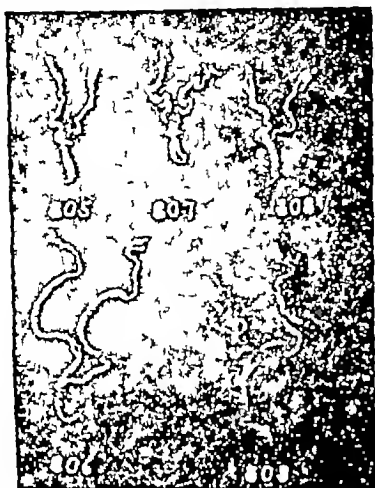


FIG 615 Effect of injection of follicular extract into eight week old rabbits Animals were litter sisters No 807 received 2 mg of extract No 805 not injected and killed at the same time as No 807 for control No 806 received 1 mg of extract daily for four days and No 809 1 mg daily for eight days No 808, uninjected control, killed on same day as No 809 (After Dowsy, Ralls Allen and Johnston)

suppression of the gonadotrophic principle of the pituitary (7) Estrogen exerts no notable effect upon the circulation, respiration, metabolic rate or skeletal growth, though the growth of other nongenital tissues may be induced by prolonged treatment, mitosis in several organs has been observed, and multiple fibromyomatous tumors sometimes develop (8) If, after menstruation has been abolished in monkeys by ovariectomy, a series of injections of estrogen is given and then stopped abruptly, menstruation occurs a few days later During the treatment of these ovariectomized animals, phenomena appear, analogous to those appearing during the estrus period of lower ani-

mals, e.g., endometrial changes, reddening and swelling of the external genitalia, and the vaginal smear shows cornified cells and the absence of leucocytes (9) A synergic action between the follicular hormone and the oxytocic principle of the posterior pituitary has been demonstrated, after the administration of the former to a mouse toward the end of pregnancy a dose of pitocin which ordinarily is ineffective causes a powerful contraction of the uterus (fig 616) (10) Estrogen acts as a "primer" for the corpus luteum hormone (progesterone), and is apparently necessary also for the maintenance of the corpus luteum during pregnancy (11) The effect of estrogen upon the nasal mucosa is described on page 895 (12) Prolonged treatment with estrogen raises the blood calcium, especially in pigeons and fowl (13) It induces

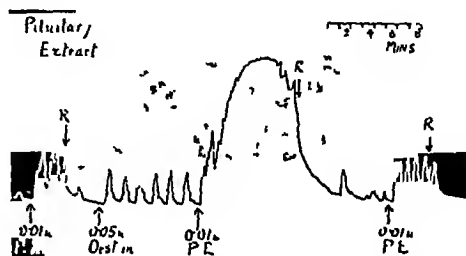


FIG 616 Action of estrin in sensitizing an isolated uterus to pitocin (After Bourne and Burn) P E, pituitary extract

water retention, an increase in blood volume and of the water content of the muscles Ovariectomy results in a loss of water and a diminished volume of blood which is restored to normal by estrogen administration

THE HORMONE OF THE CORPUS LUTEUM

After the discharge of the ovum the cavity of the ruptured Graafian follicle becomes filled by a clot of blood The small body formed in this way is sometimes spoken of as the *corpus hemorrhagicum* The clot is soon replaced by a mass of cells containing a yellow lipid material (luteal cells) These are derived from the proliferation of the epithelial cells of the membrana granulosa (*granulosa luteum cells*) and of the theca interna (*thecal luteum cells*) The follicle with its content of luteal cells constitutes the *corpus luteum* (fig 617) The circumference of the follicle by this time has become more vascular and capillaries penetrate into the yellow cell-mass The transformed follicle may now be looked upon as a temporary internal secreting

organ. If fertilization of the ovum does not occur the life of the corpus luteum is short. In the human it persists for about 10 days and then retrogresses. Its vessels become obliterated, the luteal cells disintegrate and are replaced by fibrous tissue, nothing then remains of the follicle but a pale scar—the *corpus albicans*. On the other hand, if impregnation of the ovum results the corpus luteum continues to grow and in women attains a diameter of three quarters of an inch or more by the middle of pregnancy. It then commences to

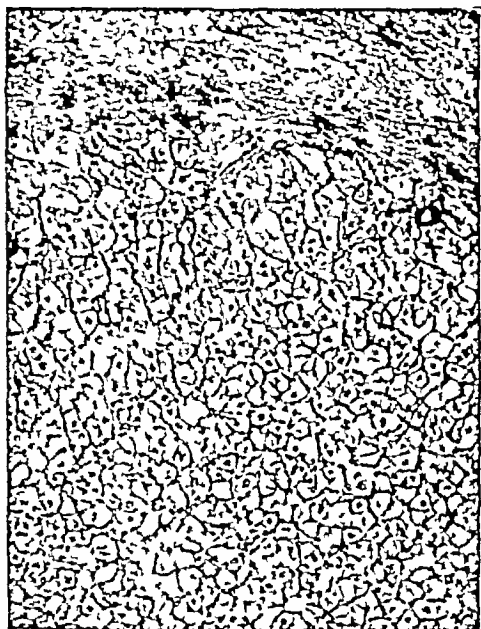
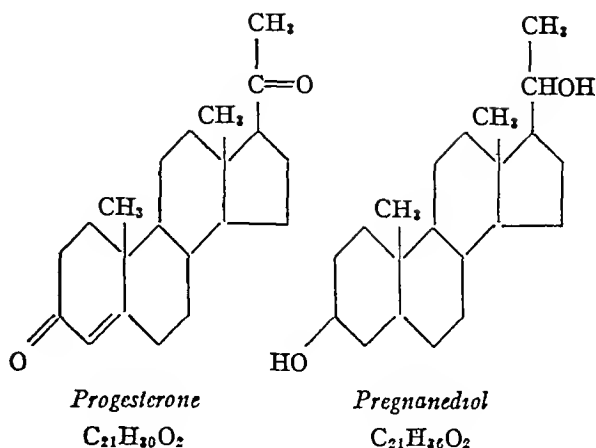


FIG 617 Section of corpus luteum showing luteal tissue under high magnification (from Parkes, *The Internal Secretions of the Ovary*, by permission of Longmans, Green & Co.)

shrink and is finally absorbed by about the seventh month.

Chemistry and functions of progesterone

The active principle of the corpus luteum has been obtained in crystalline form. Its structural formula is shown below (compare with the formula for estrone and estriol on page 878, with that for testosterone given on page 905, and with the crystalline compounds of the adrenal cortex (p 838)). This crystalline product has been named *progesterone*, the older name *progestin* being reserved for the unpurified luteal principle contained in ordinary corpus luteum extracts. Progesterone occurs in two isomeric forms, α - and β -progesterone. They are about equal in physiological activity.



Starting with a sterol (stigmasterol) obtained from soy bean Butenandt and his associates succeeded in completing the later steps in the synthesis of progesterone, and it has been prepared by these workers from pregnenediol as well as from cholesterol. Block also found that isotopically labelled cholesterol when ingested by a pregnant woman caused the appearance in the urine of pregnenediol glycuronide containing the isotope. The close relationship of progesterone to the adrenal steroids is shown by its conversion *in vitro* to desoxycorticosterone by hydroxylation at position 21.

Pregnenediol is a physiologically inert reduction derivative of progesterone isolated from the urine of pregnancy by Marrian in 1929. Its very close chemical relationship to progesterone is evident from the formulae shown above; it is readily convertible to progesterone in the laboratory. Only minute quantities of progesterone itself are found in blood or urine. Similarly to estriol and the derivatives of testosterone (p 905) pregnenediol is excreted in conjugation with glycuronic acid (Odell and Marrian). Injected progesterone appears in the urine as pregnenediol glycuronide. Its occurrence in urine in different phases of the human menstrual cycle has been investigated by Browne and associates who found it present as *sodium pregnenediol glycuronide* in the luteal phase of the menstrual cycle, but not in the follicular and intermenstrual phases. The excretion of pregnenediol commences a day or two after ovulation, reaching a maximum about a week before the onset of menstruation and ceasing 2 or 3 days before. During pregnancy much larger quantities appear in the urine, the greatest excretion being during the eighth and ninth months. In certain cases of spontaneous abortion the excretion is less than normal, a fact pointing to defective produc-

tion of the corpus luteum hormone. The liver is the chief site for the progesterone-pregnandiol conversion, as well as for the conjugation with glucuronic acid

The international standard of potency used in the assay of progesterone preparations is defined as the progestational activity of 1 mgm of the international standard crystalline preparation of progesterone. The test animal is an adult female rabbit which has been mated and then castrated, or an immature female rabbit which has been primed for 5 days previously with estrogen

The corpus luteum is essential to gestation. In some species (e.g., the rabbit and rat) it persists throughout pregnancy and abortion occurs if it is destroyed. In others (e.g., the human subject, monkey, cat and guinea pig) it is not indispensable after the earlier months of gestation (4 or 5 months in women) (see placental progesterone). It is responsible for (1) Changes in the uterine mucosa preparatory to the implantation of the ovum—pseudo-pregnancy of lower mammals and the premenstrual changes of primates. After implantation of the ovum the hormone of the corpus luteum is necessary for the development of the maternal placenta (decidua). (2) Growth of the mammary glands. (3) The suppression of estrus and ovulation.

The evidence for the foregoing is as follows: (a) Born observed that corpora lutea were not present in the ovaries of mammals which did not form a true placenta (monotremes). (b) Frankel, following up this hint, showed that if the corpora lutea were destroyed in pregnant rabbits abortion resulted. (c) If pregnant rabbits are injected toward the end of term with urine of pregnancy (p. 889), a fresh crop of corpora lutea appears and the gestation period is prolonged by several days, the fetuses become 50 per cent larger than normal and more mature (Snyder). A similar effect can be produced with the purified hormone of the corpus luteum. (d) Loeb discovered that stimulation of the uterine mucosa of the non-pregnant guinea pig by means of a glass bead or a thread during the development of the corpora lutea resulted in the growth of a small mass of decidual tissue (deciduoma) at the point of stimulation. This effect, now known as the *Loeb reaction*, could not be obtained after the corpora lutea had been excised or during a phase in the estrus cycle when they were absent. Even transplanted uterine tissue responded to stimulation in a similar way if the ovary contained corpora lutea. Other observers

have obtained corresponding results in the rat, rabbit and dog. Teel was able to produce deciduomata in the unmated rat by injections of a luteinizing extract (p. 884) of the anterior pituitary. The conclusion to be drawn from these experiments is that the contact of the fertilized ovum with the endometrium is the natural stimulus which in the presence of a corpus luteum causes the formation of decidual tissue. (e) In the rabbit which does not ovulate except after copulation a corpus luteum does not form unless this act takes place, nor does pseudo-pregnancy occur. If, however, a corpus luteum is produced by the artificial rupture of a ripe follicle, or, as first shown by Ancel and Bouin, ovulation is induced by mating with a sterile (vasectomized) male, the characteristic endometrial changes of pseudo-pregnancy appear. (f) The mammary gland, or transplanted mammary tissue, shows increased growth during pregnancy and pseudo-pregnancy which has been correlated with the growth of the corpus luteum. Mammary growth ceases at the end of pseudo-pregnancy when the corpus luteum retrogresses and is slight or absent if the latter has been prevented from forming. In rabbits, artificial rupture of ripe follicles causes corpora lutea to form in some instances but not in others, only in the former does mammary growth occur. (g) The results of several experiments indicate that the corpus luteum inhibits follicular maturation and suppresses ovulation. It is well known that squeezing the corpora lutea from the cow's ovary by manipulation through the rectum hastens the onset of the next estrous period. Destruction of the corpus luteum in the guinea-pig acts similarly. Also, a greater quantity of the follicle stimulating hormone (p. 884) of the anterior pituitary is required to induce ovulation in the presence of corpora lutea than in their absence. The function of the corpus luteum in suppressing ovulation provides against a second pregnancy being superimposed upon the first (superfetation).

The foregoing evidence for the function of the corpus luteum has been supplemented by the researches of Corner and Allen. These observers obtained an extract from the ovary of the pig which contained the active principle of the corpus luteum, thus they called *progestin*. Injections of the principle into animals produces the following effects: (a) A perfect imitation of the progestational changes (pseudo-pregnancy) in the uterine mucosa of castrated adult rabbits. The uterus of the immature rabbit or of an adult animal which has

been castrated some time previously, however, is unresponsive unless first "primed" by a previous course of estrogen injections. In order to produce the typical progestational development of the uterine mucosa in the castrated monkey it is also necessary to administer both hormones, menstruation then occurs (Smith and Engle). Kaufmann showed similarly that in castrated women a premenstrual endometrium (proliferative and secretory phases) can be produced only by the administration of both the follicular and the corpus luteum hormones. The estrogenic hormone apparently initiates the progestational changes, causing vascularization of the mucosa and proliferation of the glandular elements, while the corpus luteum stimulates secretory activity and brings about the final alterations in the mucosa necessary for the implantation and nourishment of the fertilized ovum. Estrogen also appears to be necessary for maintaining the growth of the corpus luteum of pregnancy and for the continued action of progesterin upon the uterine mucosa.

(b) Rabbits castrated early in pregnancy abort, injections of progesterin enables them to be brought to full term. (c) Softening and relaxation of the pelvic ligaments and separation of the symphysis pubis of the guinea-pig, these effects imitate those occurring in pregnancy and are not obtained in the absence of the follicular hormone. (d) Inhibition of the uterine response to pituitrin (Knaus' reaction). When the uterus of a rabbit which has been treated with progesterone is excised it does not contract when pituitrin (even in large quantities) is added to the bath in which it is suspended, definite relaxation may result. The muscle will still, however, respond to adrenaline or quinine. The motility of the uterus of the intact animal is also suppressed by progesterin. The change in reaction of the uterus of certain species to adrenaline (p 831), from inhibition to excitation, when pregnancy occurs is evidently due to the corpus luteum hormone, for a similar reversal of uterine behavior to adrenaline can be effected by progesterin injections. (e) Estrogen and progesterin are antagonistic, according to Allen 675 rat units of estrin are neutralized by 3 units of progesterin. The motility of the uterus of an infantile animal or of a castrated mature animal is inhibited by progesterone, the uterus of a mature animal with intact ovaries is affected to a much less extent by the corpus luteum hormone.

The clinical uses of progesterone are considered on p 894

Relaxin Hisaw and his associates obtained a

principle from luteal tissue which, though stated to be free from estrogen and progesterone, induced relaxation of the ligaments of the symphysis pubis of guinea-pigs. They believe that the effect is due to a separate hormone and have named it *relaxin*. This principle is also found in the blood of pregnant rabbits, and is stated to be of the nature of a peptide or polypeptide.

For some time the existence of relaxin was questioned, for it was claimed by other workers that a similar effect could be induced by estrogen or progesterone, or a combination of the two. Abramowitz and his associates found, however, that relaxin prepared from sows' ovaries administered in a $\frac{1}{100}$ dosage as great as that required to cause an estrogenic or progestational effect

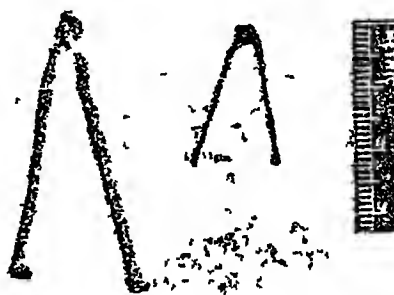


FIG 61.8 Effect of hypophysectomy upon the uterus. On right, uterus of normal rat, on left, that of an hypophysectomized rat. (After Van Dyke, *The Physiology and Pharmacology of the Pituitary Body*, Chicago Univ. Press.)

caused relaxation of the pubic ligaments. It was found further by Hisaw and his associates that after hysterectomy progesterone was inert insofar as relaxation of the ligaments and mobility of the symphysis were concerned, whereas relaxin was still effective. The response to relaxin is also much more prompt than that to progesterone. These facts leave little doubt of the existence of relaxin, and have led this group of workers to believe that progesterone causes the production or the release of relaxin from the uterus.

THE GONADOTROPHIC (GONAD-STIMULATING) HORMONES OF THE ANTERIOR LOBE OF THE PITUITARY, GONADOTROPHINS

Several observations in the past have pointed to the anterior pituitary as exerting an influence upon sexual development. The gradual atrophy and suppression of the sex functions in diseases of the anterior lobe in man (acromegaly, Frohlich's syndrome) and the atrophy of the gonads after hypophysectomy in animals are among some of these observations (fig 61.8).

In the earlier experiments of Evans and associates with a crude saline growth-promoting extract

of the anterior lobe of the pituitary (p 785) it was observed that the Graafian follicles became enlarged and filled with luteal cells (fig 61 9) Usually,

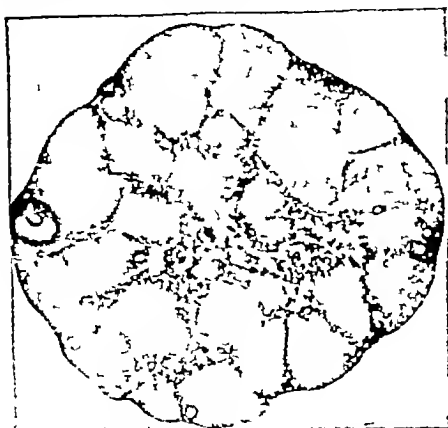


FIG 61 9 Ovary of adult mouse injected with alkaline extract of anterior lobe of pituitary. Large numbers of corpora lutea are present, and few follicles (from Parkes, *The Internal Secretions of the Ovary*, by permission of Longmans, Green & Co.)

remained fertile. Ovulation in hens was inhibited. These effects upon the sex processes were shown later to be independent of the growth hormone, for it was possible to separate the extract into two fractions, one having an effect upon growth, the other having the luteinizing effect just described.

It soon appeared that another sex hormone possessing a quite different effect was present in the anterior lobe. Smith and Engle found that when *fresh* mammalian anterior lobe tissue was transplanted daily into immature female rats or mice, estrus was precipitated and ovulation stimulated. The ovaries enlarged to 10 times the normal size (fig 61 10), and developed a large number of follicles which ripened and discharged a "shower" of ova (superovulation). Corpora lutea formed within the ruptured follicles, reached a small size and then retrogressed. Under the influence of the transplants the vagina, which is not a complete canal in the immature rat or mouse, opens, its epithelium changes from the columnar to the squamous type and the uterus enlarges from 5 to



FIG 61 10 Showing effect of anterior pituitary upon ovaries. On left, below, ovaries of rat after twelve implantations of fresh rat pituitary gland; above, ovaries of litter mate control rat (After Collip). On right, below, follicular maturation induced in immature rat on twenty-ninth day by eight daily implantations of fresh pituitary gland; above, ovary of litter mate control. (After Smith and Engle.)



of course, corpora lutea form only after ovulation, but in these animals the estrus cycles were suppressed, and the ova remain imprisoned within the *luteinized* follicles. The extract was later shown to depress the sex instincts of males and to reduce the weight of the testes. The male animals, however,

10 times. Ovulation and estrus were also induced in adult sexually quiescent females. The testes, seminal vesicles and penis of immature males were stimulated to increased growth, and gonadal atrophy which follows hypophysectomy in either males or females was prevented by the anterior lobe

transplants Zondek and Aschheim about the same time performed similar experiments in Germany upon immature female mice and obtained comparable results, the effects being evident within 100 hours. They pointed out that the ovaries must be intact in order for estrus to occur. The essential action upon the female of this anterior

extraction of the anterior lobe. The follicle-stimulating hormone (whether in anterior pituitary or in pregnancy urine) was called *prolan A* and the luteinizing hormone *prolan B* (terms rarely used today). The gonadotrophic principles were later named the *follicle-stimulating* and the *luteinizing hormones*, respectively.

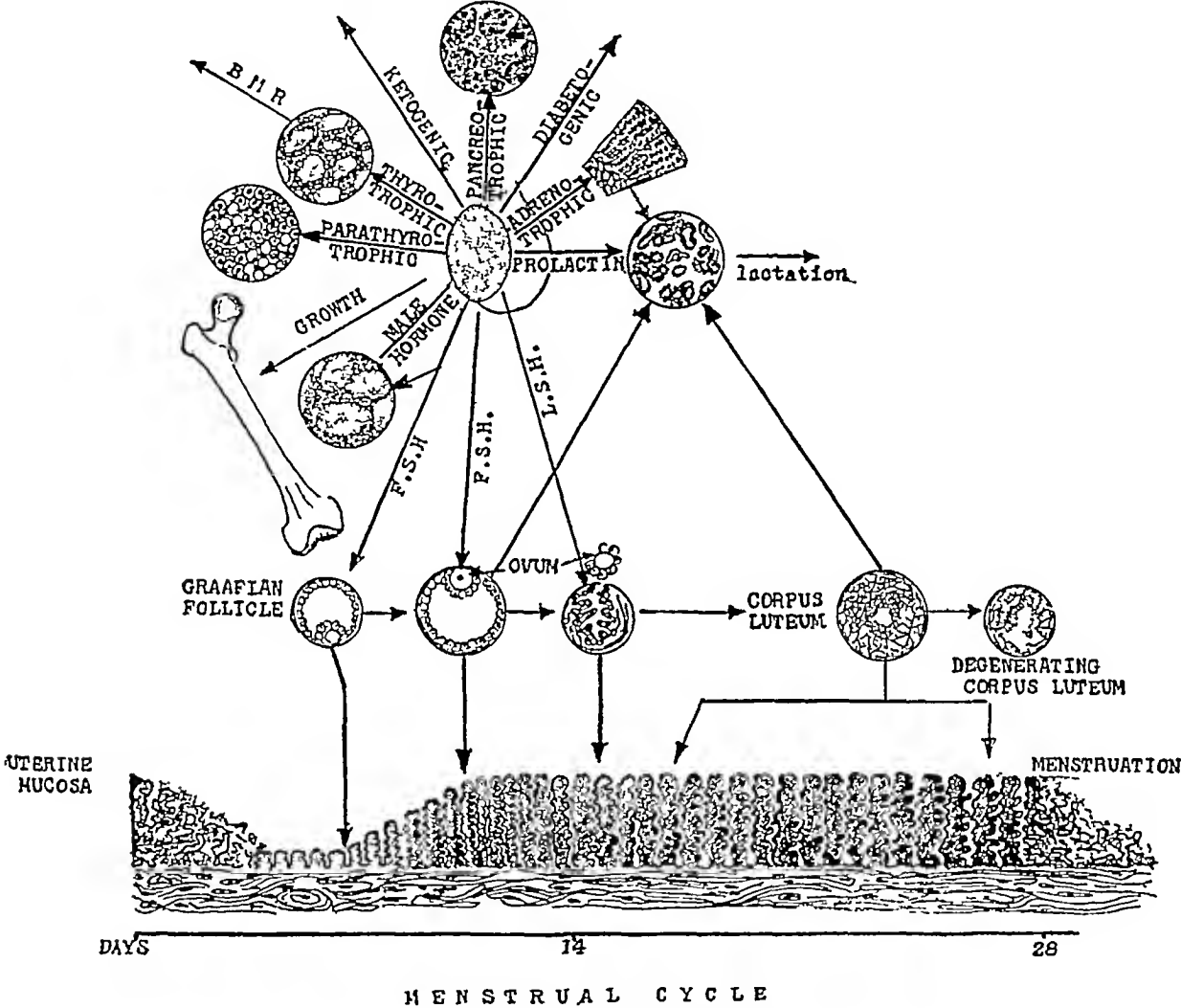


FIG 61 11 Diagram summarizing the endocrine relationships of the anterior lobe of the pituitary

lobe hormone is therefore to *stimulate follicular maturation and cause the liberation of the estrogenic hormone*. Estrus and its associated phenomena are secondary to the action of estrogen. Aschheim and Zondek also showed that both types of effect, luteinization on the one hand, and follicular maturation (ripening) on the other, could be produced in female mice by injections of the blood or urine of pregnant women (see pregnancy test below). Later work has shown that the follicle-stimulating principle may be obtained by acid

The follicle-stimulating hormone also stimulates the tissue of the testes analogous to the granulosa of the ovarian follicle, namely, the epithelium of the seminiferous tubules, large numbers of sex cells in various stages of development, including mature spermatozoa, are produced. The luteinizing principle, on the other hand, acts apparently upon the interstitial cells of the testes which are analogous to the theca interna cells of the ovary (p 874), causing the liberation of the testicular hor-

mone¹⁰ For this reason interstitial cell stimulating hormone (ICSH) has been suggested as a more suitable term than luteinizing hormone (see diagram, fig 61 11) Thus, through the discovery of these gonadotrophic hormones the arrest of sexual development in young animals and the atrophy of the gonads in adults after hypophysectomy is explained

The respective actions of the gonadotrophic hormones of the anterior pituitary are tabulated below

Follicle stimulating hormone ¹¹ (FSH)	Ripening of follicles, superovulation, small corpora lutea.
	Production of estrogen, phenomena of estrus
	Proliferation of epithelium of the seminiferous tubules of the testes.
Luteinizing hormone (LH, ICSH)	Extensive luteinization, retained ova.
	Production of hormone of corpus luteum.
	Inhibition of estrogen production suppression of estrus.
	Stimulation of interstitial tissue of testes (p 906)

These two hormones act synergically, the luteinizing principle being without effect unless preceded by the action of the follicle-stimulating hormone. The luteinizing principle is therefore inactive after hypophysectomy unless the animal has been primed with an estrogen

The maintenance of the corpus luteum in a functioning state after it has been formed does not depend upon the luteinizing hormone (ICSH) as was originally thought, but upon the estrogenic hormone and a pituitary principle which has been called the *luteotrophic* principle. Purified LH administered to rats fails to maintain the luteal tissue, whereas, purified lactogenic hormone (prolactin, p 897) is effective. It is probable, therefore, that the luteotrophic principle and prolactin are identical.

¹⁰ Some workers believe that the gonadotrophic functions of the pituitary are due to a single hormone and that the type of response obtained is dependent upon the dosage or upon the stage of the sexual cycle during which the treatment is given. Wallen Lawrence and Fevold and Hisaw, and a number of other workers have secured evidence strongly favoring the belief in two distinct hormones. They have obtained two nearly chemically pure preparations the one having a follicle-stimulating action alone, the other causing luteinization. The question of one or two hormones remains unsettled

¹¹ Smith suggests that since this principle stimulates the germinal tissue of both sexes it be called the *gonadotrophic hormone*

Factors causing rupture of the ripe Graafian follicle—ovulation

Most of the experimental work upon this question has been performed upon rabbits, though some observations have also been made upon monkey's. It was shown by Bellerby that ovulation, which does not occur spontaneously in the rabbit (p 876), may be induced in this animal by injecting an acid extract of the anterior lobe. On the other hand, hypophysectomy of the rabbit up to one hour after copulation prevents ovulation, which normally occurs 10 to 12 hours following the act (Fee and Parkes) Hypophysectomy performed later than one hour after copulation does not inhibit ovulation These observations imply that rupture of the follicle and discharge of the ovum under ordinary circumstances is, in this animal, the result of a hormonal influence of some sort exerted by the pituitary The pituitary activity responsible for the effect apparently occurs within one hour after copulation It is not generally believed that any specific "ovulation hormone" exists, but that ovulation is the result of a balance struck between the follicle-stimulating and the luteinizing hormones. Afferent impulses from the genital tract are apparently responsible, in part, for calling into play the pituitary response which leads to ovulation, but they are not essential, for Fee and Parks have shown that ovulation occurs after the vulva and vagina have been anesthetized Sexual excitement resulting in an orgasm is capable alone (i.e., without coitus) of causing ovulation in the rabbit. That the influence upon the pituitary in this instance is of central origin is indicated by the observation that ovulation may be induced in the rabbit by electrical stimulation of the cerebrum or lumbosacral cord (Marshall and Verney), by the intravenous injection of the convulsant drug picrotoxin or of copper acetate and certain other chemicals. Ovulation can also be induced by stimulation of the cervical sympathetic, the hypothalamus, the hypothalamico-hypophyseal tract of nerve fibers or of the pituitary itself

The ovulation response is abolished with difficulty, for impulses reaching the pituitary from many sources are capable of inducing it. Brooks found that it was not prevented after coitus though the sacral spinal cord, both sympathetic chains and the uterus, together with the proximal half of the vagina had been removed. Nor is it abolished by ablation of the neocortex and olfactory lobes. It no longer occurs, however, after section of the pituitary stalk. It is also abolished in animals whose hind limbs have been paralyzed by section of the lumbar spinal cord Sexual excitement aroused by proprioceptive and tactile impulses from the limbs as the animal accommodates its posture for the act of copulation appears to be an indispensable factor in the response. The conclusion to be drawn from these various types of experiment is that though genital stimulations may elicit the response, they are not

essential. On the other hand, the cooperation of the animal in the act of coitus and the associated sexual excitement appear to be necessary factors.

According to Markee and associates the effect of nervous excitement in inducing ovulation is mediated by adrenaline, that is, nervous impulses cause the liberation of the latter hormone which then acts directly upon the secretory cells of the anterior lobe. Adrenaline applied directly to the latter was followed by ovulation.

The following is a summary of the pituitary influence upon the sex mechanism. The follicle-stimulating principle initiates sexual activity. The ovaries or testes, stimulated by this anterior lobe hormone, bring about the psychic phenomena of puberty and of the estrus cycles of post-pubertal life, as well as the physical changes in the accessory organs (vagina, uterus and seminal vesicles) characteristic of the sexually mature animal. It stimulates the development of the ova and granulosa cells of the follicles and, in the male, the epithelium of the seminiferous tubules. The luteinizing hormone depresses the first part of the ovarian cycle and encourages the second or luteal phase, thereby bringing about the uterine changes associated with pseudo pregnancy or with gestation. It acts specifically upon the cells of the theca and granulosa with the production of progesterin, or, *in the case of the male*, upon the interstitial cells of the testes with the production of male hormone.

The basophil cells (p. 783) are probably the source of the gonad-stimulating hormones, after castration the basophil elements of the anterior pituitary of the rat are increased in number and size ("castration cells"¹²) while the gland's content in gonadotrophic principles increases. Moreover, when a castrated male rat and an hypophysectomized female are united parabiotically the female enters into continuous estrus and its ovaries show large numbers of follicles (superovulation). These results indicate that the relationship between the pituitary and the gonad is not one-sided but that the latter exerts, normally, a restraining influence upon the gonadotrophic functions of the former. Such a relationship is supported by the converse observation, that continued treatment with estrogen reduces the gonad-stimulating power of the anterior pituitary, the ovaries showing atrophic changes. The changes in the anterior lobe induced by estrogen injections are, increased vascularity,

reduction in number and ultimate disappearance of basophil and acidophil cells with a proportionate increase in chromophobes. It is considered probable that the sequence of the two phases of the ovarian cycle is governed through the interplay of ovarian and pituitary hormones. When, for example, as a result of follicular stimulation the concentration of estrogenic hormone in the blood reaches a certain level the liberation of the follicle-stimulating hormone of the anterior pituitary is suppressed and discharge of luteinizing hormone brought about.

The excretion of gonadotrophic hormones in the urine

It has been assumed that the gonadotrophic effects of human pregnancy urine discovered by Aschheim and Zondek were due to the gonadotrophins of the pituitary. These it was supposed were elaborated in larger than usual amounts during pregnancy, the excess being excreted by the kidney. But it is now generally agreed that the gonadotrophic effects of pregnancy urine are not caused by the pituitary hormones, they depend upon a *single* principle of a different nature (see placental hormones below). Collip and his associates showed that this gonadotrophic principle was identical with that of the placenta, and named it the *anterior-pituitary like substance* (APL) which is described below. It is now most commonly referred to as the *human chorionic gonadotrophin* (HCG).

On the other hand, a gonadotrophin believed to be the follicle stimulating hormone of the pituitary is excreted in the urine of castrated animals and men, and in the urine of women after the menopause or following ovariectomy. The appearance of this pituitary hormone in the urine may also be associated with *new growths* of the generative organs (both benign and malignant but especially the latter), and with *raised intracranial pressure*. Occasionally it is found in the urine of normal men or of non-pregnant women before the menopause. This principle, unlike HCG, is effective when tested upon normal monkeys or hypophysectomized rats.

PLACENTAL HORMONES

The human placenta contains (a) the two forms of estrin, *estrone* (ketohydroxyestrin or theelin, p. 877) and *estriol* (trihydroxyestrin or theelol), (b) *progesterone*, and (c) the *anterior-pituitary-like substance* (APL), or *human chorionic gonadotrophin* (HCG). Collip obtained the placental gonadotrophin by acetone extraction and showed that though it had an action simulating the combined effects of

¹² These structures contain a large vacuole, the latter by pushing the nucleus toward the circumference of the cell gives it an appearance resembling that of a signet ring.

FSH and LH (ICTH) it is a single substance differing in several respects from any known pituitary principle. It was therefore termed "anterior-pituitary-like (APL)." From evidence derived from tissue culture and histochemical experiments, human chorionic gonadotrophin (HCG) is generally believed to be secreted by the cytotrophoblast (Langhans layer) of the developing embryo.

SUMMARY OF SOME OF THE MORE IMPORTANT OBSERVATIONS RELATING TO THE NATURE OF HUMAN CHORIONIC GONADOTROPHIN

(1) When immature female rats are treated with chorionic gonadotrophin prepared from human placenta, placental implants, or with extracts of human pregnancy urine, follicular maturation, estrus, and the formation of corpora lutea result, the luteinizing action is the predominant effect. ICSH is ineffective in immature animals. The failure of the pituitary principle in this respect is believed to be due to the absence of FSH whose synergic action is required by both hormones, but whereas the placental principle appears to actually stimulate the production of FSH the pituitary gonadotrophin (ICSH) does not. Human chorionic gonadotrophin causes enlargement of the ovaries, but to a less degree than does the pituitary principle, and hypertrophy of the interstitial cells (cells of Leydig) of the testes, as a consequence, the output of male hormone is increased, and the growth of the accessory sex organs stimulated. The latter action is of value in the treatment of undescended testes and of eunuchoidism. Chorionic gonadotrophin will also maintain the integrity of the seminiferous tubules in hypophysectomized animals, but unlike the pituitary principle will not prevent ovarian atrophy after hypophysectomy, for it requires for this effect the synergic action of FSH. In contrast to ICSH the placental principle exerts little or no effect upon the growth of the immature testes of birds.

(2) Though extractable in rather large amounts from human placental tissue the anterior pituitary-like principle is absent from the placentae of lower mammals, it is present in the urine of pregnant apes, it disappears rather promptly from the urine of women after parturition.

(3) In chorionepithelioma, a malignant tumor of placental (chorionic) tissue, and hydatidiform mole, a cystic degenerative disease of chorionic tissue, HCG is present in large amounts in the pathological tissue and in high concentration in the urine. The gonadotrophic hormone is found in high concentration in the chorionic villi of even very young embryos. It also appears in the urine of males suffering from tumors containing chorionic tissue.

(4) The pituitaries of animals or of women do not contain more gonadotrophin in the pregnant than in the non-pregnant state, a finding inconsistent with an

increased production of the pituitary principle during gestation, and a spilling over of the hormone through the kidney.

(5) The placental and pituitary principles are differentiated also by the results of repeated injections. Animals which have become resistant to the action of HCG will still respond to anterior lobe extracts, and conversely those which have become tolerant to the latter will still respond to HCG.

The foregoing summary indicates that the substance of the placenta and the gonadotrophic principle of pregnancy urine are identical, i.e., the gonadotrophic principle in the urine of pregnancy is derived, not from the pituitary but from the placental (chorionic) tissue. A possible function of the gonadotrophin of the placenta is that it serves to supplement the action of the pituitary in maintaining the growth of the corpus luteum during pregnancy. The low concentration of gonadotrophin in the pituitary of pregnancy conforms with such a possibility. According to this idea the placenta furnishes a hormone which, acting through the mediation of the ovary, ensures its own physiological activity.

The Aschheim-Zondek test for pregnancy

The experiments of Aschheim and Zondek have led to the development of a practical test for pregnancy. After the injection of 2 cc of urine of a pregnant woman (divided into 6 doses given over a 2-day period) into an immature female mouse 3 to 4 weeks old, the following ovarian changes are produced:

(1) Ripening of follicles and estrus

(2) Hemorrhages into some unruptured follicles, giving rise to "blood spots" (Blutpunkte) about the size of a pin's head.

(3) Luteinization of follicles in which the ova are retained—atretic corpora lutea.

Effects (2) and (3) appear usually within 100 hours and permit a diagnosis of pregnancy to be made. Reaction (1) is not entirely specific since it occurs in conditions other than pregnancy, e.g., after the menopause (p. 893) and in new growths of the genital organs (see p. 889), especially carcinoma. The test gives a correct result (positive or negative) in nearly 99 per cent of trials. A diagnosis of pregnancy as early as the fifth day after the first missed menstrual period can be made by the use of this test.¹³ The test fails in lower animals

¹³ Estrus does not appear in large amounts in the urine until a much later date and small amounts are present in the urine of non-pregnant women. Aschheim states

(e.g., rat, dog, cat, etc.) and the monkey, but is positive in the chimpanzee

The test depends upon the presence of living placental (chorionic) tissue and the production of HCG which acts synergically with FSH liberated by the pituitary. Consequently a positive result is obtained if certain abnormalities of gestation such as tubal pregnancy, hydatidiform mole or chorionepithelioma exist. In the latter two conditions unusually large amounts of the gonadotropic principle are present in the urine. The test is also positive in threatened abortion but becomes negative when detachment and death of the ovum occur. Metastatic tumors of chorionepithelioma cause a positive reaction after removal of the uterus, the test may therefore prove invaluable as an aid in the diagnosis of the dissemination of this malignant disease. In the male, testicular tumors composed of malignant embryonal tissue (teratoma, epithelioma) also cause a positive test, such growths may result in the appearance of relatively enormous amounts of gonadotropic substance in the urine. Many cases of pituitary tumor give a positive test.

Friedman has modified the foregoing technique by substituting rabbits for mice as test animals. Since rabbits, ordinarily, ovulate only after sexual excitement, adult animals may be used for the test. This is a distinct advantage, as is also the fact that the ovarian reaction is developed much earlier—within from 16 to 36 hours. The animals must have been isolated from the males for at least three weeks prior to their use in order to ensure that natural ovulation has not occurred. Five to 10 cc. of pregnancy urine are given in a single intravenous injection. The percentage of correct diagnoses is about the same as when immature mice are employed.

Placental progesterone

Though the production of progesterone by the placenta, as distinct from mere storage, has not been proved conclusively, the presumptive evidence of such a function is so strong that there is little room for doubt. Progesterin is thought to be secreted by the syncytial layer of the placenta, though attempts to demonstrate this by growing

that the diagnosis of pregnancy by means of the urine is a very ancient practice. In an Egyptian papyrus some 3000 or 4000 years old it is directed that should a woman wish to know whether or not she is pregnant she should place some earth and barley in a vessel and add a little of her urine each day. Should the barley grow she is pregnant. It may be remarked that estrin, which as already mentioned, is present in high concentration in the urine after the first month or two of pregnancy, is a stimulant to plant growth.

placental tissue *in vitro* have failed. It has been calculated that approximately one-third of the progesterone secreted during pregnancy is derived from other than ovarian tissue. Most of this is believed to be of placental origin, a small part presumably is furnished by the adrenal cortex.

The following is a summary of the presumptive evidence for a placental progesterone: (1) In man and some other mammals the ovaries can be excised in the later part of pregnancy (at a time when the corpus luteum normally undergoes regression) without interrupting pregnancy. (2) If in the rat whose corpus luteum persists naturally until the end of gestation, all the fetuses but one are removed, but all placentae retained, pregnancy continues without interruption after ovariectomy (Haterius). (3) Collip and his associates also found that in the rat ovariectomy and the removal of all fetuses, but leaving the placentae undisturbed, did not arrest the endometrial development characteristic of the pregnant state. (4) In the cat ovariectomy early in pregnancy causes the death of the embryos but the placentae continue to develop and the progestational activity of the endometrium progresses.

These results prove that the embryo itself does not produce a hormone which directly or indirectly, through the maintenance of luteal tissue, is essential for the endometrial changes of pregnancy.

THE MENSTRUAL CYCLE

UTERINE CHANGES

In primates (man, anthropoid apes and higher monkeys) the reproductive organs pass through a series of changes at periodic intervals, known as the menstrual cycle. The most evident phenomenon of the cycle is the escape of blood from the vagina. The hemorrhage has its origin in the endometrium and is called *menstruation*.

Markee has followed the process of menstruation microscopically in patches of endometrium transplanted into the anterior chamber of the eyes of monkeys. Bleeding occurred from the transplanted tissue during menstruation, but preceded the appearance of blood in the vagina by about 3 hours. Estrogen injections produced dilatation of the vessels of the graft. These observations emphasize the essentially endocrine nature of the uterine changes occurring during the menstrual cycle. In the human subject patches or tumors of endometrial tissue are sometimes found in extra-uterine situations, e.g., surface of the broad ligament or ovary, in the omentum, pelvic peritoneum or subcutaneous tissue of the vulva or perineum, or in the tissue in the

neighborhood of a laparotomy scar The ectopic tissue bleeds during menstruation The condition is not very uncommon It is referred to as *endometriosis*

It is generally believed that rhythmical variations in the activity of the anterior pituitary are primarily responsible for the regular recurrence of the menstrual cycle, the ovary (through the discharge of estrogen and progesterone) playing the rôle of intermediary Menstruation is the most obvious event in the menstrual cycle and is customarily described as the first stage, but actually it is the culminating stage It will therefore be placed last in the following division of the cycle This order is also more convenient in correlating the ovarian cycle with the uterine changes

(1) The *stage of repair and proliferation—follicular phase*¹⁴ During this stage which lasts, on the average, for 15 days counting from the first day of the menstrual bleeding the epithelium of the endometrium which was shed during the menstrual flow ((3) below) is restored The uterus enlarges as a result of the growth of its stroma, it becomes more vascular, its arteries become coiled, the epithelial lining hypertrophies and the glands show proliferative changes Ovulation occurs at about the end of the proliferative stage which therefore corresponds to pro-estrus of animals These uterine changes are dependent upon the action of the estrogenic hormone

(2) The *premenstrual or secretory stage—luteal phase* This stage, which follows the follicular phase, lasts for 13 or 14 days The uterine mucosa shows marked hypertrophy and is highly vascular, its glands become elongated and assume a coiled or corkscrew form The glandular secretion becomes greatly increased and more mucoid in character This stage is dependent upon the action of the corpus luteum, but also, apparently, upon the presence of estrogen which continues to be secreted, it corresponds to pseudo-pregnancy of certain animals Toward the end of the premenstrual stage the endometrium resembles the decidua of early pregnancy, typical decidual cells appearing in the uterine stroma Toward the end of the luteal phase the vessels of the mucosa constrict. Swelling of the mammary glands and often mild psychic disturbances (irritability, nervousness, depression, etc.) occur This stage may be absent, though rarely, from the menstrual cycle (p 892)

¹⁴ Some divide the stage into two—a period of regeneration and restoration of the shed epithelium followed by one of proliferation.

(3) The *destructive stage or stage of menstrual flow* lasts for about four days The vasoconstriction toward the end of the luteal phase and resulting ischemia of the endometrium lead to necrosis of the superficial layers, dilatation of the vessels then ensues accompanied by shedding of the necrosed tissue and bleeding from the denuded surface.

Menstrual blood as it appears externally is incoagulable This is probably due to its containing a substance (fibrinolysin) which has destroyed the fibrin in the clots previously formed in the uterus or vagina

OVARIAN CHANGES

The uterine changes are associated with just as definite changes in the ovary During the stage of proliferation of the endometrium the Graafian follicle is undergoing maturation Ovulation most usually occurs around the 15th day (13th to 17th) after the first day of the last menstruation, that is, about mid-way between two menstrual bleedings or "periods" and, as already mentioned, at about the end of the proliferative stage Corner has recovered unfertilized ova at this time from the Fallopian tubes of monkeys, and Newell, Allen, Pratt and Bland, from the tubes of women During the endometrial hyperplasia of the premenstrual stage the corpus luteum is developing It reaches its maximal size at the end of this period if fertilization of the ovum does not occur, and its subsequent degeneration coincides with the onset of the menstrual flow (figs 61 11 and 61 12) If fertilization and successful implantation of the ovum occur the corpus luteum continues to enlarge (p 882) The excision of a recently formed corpus luteum is followed by menstruation or, if implantation of the ovum has occurred, by abortion

Ovulation is accompanied by a sharp rise in electrical potential between the uterine cervix and the abdominal wall This phenomenon was first demonstrated in the rabbit by Burr and his associates in 1935 and has since been fully confirmed by others The potential change amounts to from 6 to 10 microvolts and lasts for about an hour It is detected by means of electrodes placed in the situations mentioned, and connected to a sensitive potentiometer which activates a moving-cell galvanometer The potential change can be recorded photographically This method has been employed to ascertain the time of ovulation in women, the electrical change having been correlated with the presence of a recently ruptured follicle in the ovary

The *body temperature*, recorded daily during the menstrual cycle and plotted, shows an abrupt dip, usually in the mid-menstrual period (13th to 17th day) The curve rises sharply again and reaches a level about 0.8° Fahrenheit above normal, which it maintains until the commencement of the menstrual flow when it falls back to normal

Ovulation is also followed by the appearance of pregnanediol in the urine (p 881) The detection of this derivative of progesterone is therefore taken as evidence of ovulation and fixes approximately the time of its occurrence

Two maxima occur in estrogen excretion, the first at the time of ovulation, the second at the height of the development of the corpus luteum, a sharp fall occurs just before the commencement of menstrual bleeding

Information as to the time of ovulation in women by various methods has a practical value in contraception The so-called "safe period", i.e., during which conception presumably cannot occur, has been placed in those parts of the cycle before and after ovulation But the time of ovulation varies considerably in different women and may occur several days before or after the 15th day of the cycle Pregnancy has followed artificial insemination as early as the 4th or 5th day of the cycle It cannot be said, therefore, what is the "safe period" in any individual case, but the least likely time for conception to occur is within the last 8 days or so of the cycle, that is, 8 days before menstruation

It would seem that the menstrual rhythm once initiated is perpetuated automatically through the interaction of the pituitary and ovary From the facts in hand it is possible to draw tentatively the following picture of the governing mechanism Estrin (estradiol) increases in the body fluids, due to stimulation of the ovary by the follicle-stimulating hormone of the pituitary, but when the concentration of this ovarian hormone reaches a certain level it acts in turn to suppress the output of the follicle-stimulating hormone, the concentration of estrogen upon which the proliferative stage of menstruation depends is thus reduced The hypophysis now releases its luteinizing principle which stimulates the development of the corpus luteum, but as the concentration of progesterone rises the production of the luteinizing principle is in turn suppressed, with the result that the integrity of the luteal tissue cannot be maintained The concentration of estrogen having fallen by this time to a low value, menstrual bleeding occurs and the

secretion of follicle-stimulating hormone is then resumed—another cycle commences

THE FACTORS CONCERNED IN THE FERTILIZATION OF THE OVUM

The ovum if unfertilized degenerates within a few hours after its discharge from the ovary, probably in the Fallopian tube For this reason, and because only one ovum is discharged in each cycle, the period during which fertilization can occur is very short, probably not longer than 6 or 7 hours Yet the period during which coitus may lead to conception is relatively long, for the spermatozoa retain their power to fertilize the ovum for from 4 to 5 days If, for example, the sperm is deposited say on the 11th day of the cycle and ovulation occurs on the 15th, it is possible for the ovum to be fertilized The sperm shows motility for some time after it has lost its ability to fertilize The former property is therefore not a reliable index of the latter

As mentioned in the preceding section, the sperm must penetrate the ovum within a few hours after ovulation or cannot do so at all Fertilization is effected in the Fallopian (uterine) tube and here also the earlier stages in the maturation of the ovum take place When discharged, the ovum is surrounded by the *cumulus oophorus* It has been thought that this covering could not be penetrated by the sperm, and must first be destroyed by the enzyme hyaluronidase contained in the sperm of most mammalian species But there has been a misconception in this regard, for the male sex cell can reach the ovum through an intact cumulus It is possible, however, that the enzyme aids the sperm in some other way in its approach and penetration of the ovum An hyaluronidase inhibitor, such as tri-gentisic acid (rehibin) added to rabbit semen before insemination, was found by Parkes to prevent fertilization—an effect not due merely to a non-specific spermicidal action However, the administration of the hyaluronidase inhibitor to the animals was without effect upon fertility¹⁵

A CONSIDERATION OF THE FACTORS CONCERNED IN THE PRODUCTION OF THE MENSTRUAL FLOW

Hartman's studies have shown beyond doubt that menstruation occurs without ovulation in the

¹⁵ Martin and Bieler reported that the hyaluronidase inhibitor phosphorylated hesperidin administered to rats diminished fertility, Sieve later made the arresting announcement that this compound was anti-contraceptive in the human subject Chang and Pincus have been unable, however, to confirm these *in vivo* findings

neighborhood of a laparotomy scar. The ectopic tissue bleeds during menstruation. The condition is not very uncommon. It is referred to as *endometriosis*.

It is generally believed that rhythmical variations in the activity of the anterior pituitary are primarily responsible for the regular recurrence of the menstrual cycle, the ovary (through the discharge of estrogen and progesterone) playing the rôle of intermediary. Menstruation is the most obvious event in the menstrual cycle and is customarily described as the first stage, but actually it is the culminating stage. It will therefore be placed last in the following division of the cycle. This order is also more convenient in correlating the ovarian cycle with the uterine changes.

(1) *The stage of repair and proliferative or—follicular phase.*¹¹ During this stage which lasts, on the average, for 15 days counting from the first day of the menstrual bleeding the epithelium of the endometrium which was shed during the menstrual flow ((3) below) is restored. The uterus enlarges as a result of the growth of its stroma, it becomes more vascular, its arteries become coiled, the epithelial lining hypertrophies and the glands show proliferative changes. Ovulation occurs at about the end of the proliferative stage which therefore corresponds to pro-estrus of animals. These uterine changes are dependent upon the action of the estrogenic hormone.

(2) *The premenstrual or secretory stage—luteal phase.* This stage, which follows the follicular phase, lasts for 13 or 14 days. The uterine mucosa shows marked hypertrophy and is highly vascular, its glands become elongated and assume a coiled or coiled-screw form. The glandular secretion becomes greatly increased and more mucoid in character. This stage is dependent upon the action of the corpus luteum, but also, apparently, upon the presence of estrogen which continues to be secreted, it corresponds to pseudo-pregnancy of certain animals. Toward the end of the premenstrual stage the endometrium resembles the decidua of early pregnancy, typical decidual cells appearing in the uterine stroma. Toward the end of the luteal phase the vessels of the mucosa constrict. Swelling of the mammary glands and often mild psychic disturbances (irritability, nervousness, depression, etc.) occur. This stage may be absent, though rarely, from the menstrual cycle (p. 892).

¹¹Some divide the stage into two—a period of regeneration and restoration of the shed epithelium followed by one of proliferation.

(3) *The destructive stage or stage of menstrual flow* lasts for about four days. The vasoconstriction toward the end of the luteal phase and resulting ischemia of the endometrium lead to necrosis of the superficial layers, dilatation of the vessels then ensues accompanied by shedding of the necrosed tissue and bleeding from the denuded surface.

Menstrual blood as it appears externally is incoagulable. This is probably due to its containing a substance (fibrinolysin) which has destroyed the fibrin in the clots previously formed in the uterus or vagina.

OVARIAN CHANGES

The uterine changes are associated with just as definite changes in the ovary. During the stage of proliferation of the endometrium the Graafian follicle is undergoing maturation. Ovulation most usually occurs around the 15th day (13th to 17th) after the first day of the last menstruation, that is, about mid-way between two menstrual bleedings or "periods" and, as already mentioned, at about the end of the proliferative stage. Corner has recovered unfertilized ova at this time from the Fallopian tubes of monkeys, and Nevell, Allen, Pratt and Bland, from the tubes of women. During the endometrial hyperplasia of the premenstrual stage the corpus luteum is developing. It reaches its maximal size at the end of this period if fertilization of the ovum does not occur, and its subsequent degeneration coincides with the onset of the menstrual flow (figs. 61.11 and 61.12). If fertilization and successful implantation of the ovum occur the corpus luteum continues to enlarge (p. 882). The excision of a recently formed corpus luteum is followed by menstruation or, if implantation of the ovum has occurred, by abortion.

Ovulation is accompanied by a sharp rise in electrical potential between the uterine cervix and the abdominal wall. This phenomenon was first demonstrated in the rabbit by Burr and his associates in 1935 and has since been fully confirmed by others. The potential change amounts to from 6 to 10 microvolts and lasts for about an hour. It is detected by means of electrodes placed in the situations mentioned, and connected to a sensitive potentiometer which activates a moving-cell galvanometer. The potential change can be recorded photographically. This method has been employed to ascertain the time of ovulation in women, the electrical change having been correlated with the presence of a recently ruptured follicle in the ovary.

menstrual cycle and plotted, shows an abrupt dip, usually in the mid-menstrual period (13th to 17th day) The curve rises sharply again and reaches a level about 0.8° Fahrenheit above normal, which it maintains until the commencement of the menstrual flow when it falls back to normal

Ovulation is also followed by the appearance of pregnanediol in the urine (p 881) The detection of this derivative of progesterone is therefore taken as evidence of ovulation and fixes approximately the time of its occurrence

Two maxima occur in estrogen excretion, the first at the time of ovulation, the second at the height of the development of the corpus luteum, a sharp fall occurs just before the commencement of menstrual bleeding

Information as to the time of ovulation in women by various methods has a practical value in contraception The so-called "safe period", i.e., during which conception presumably cannot occur, has been placed in those parts of the cycle before and after ovulation But the time of ovulation varies considerably in different women and may occur several days before or after the 15th day of the cycle Pregnancy has followed artificial insemination as early as the 4th or 5th day of the cycle It cannot be said, therefore, what is the "safe period" in any individual case, but the least likely time for conception to occur is within the last 8 days or so of the cycle, that is, 8 days before menstruation

It would seem that the menstrual rhythm once initiated is perpetuated automatically through the interaction of the pituitary and ovary From the facts in hand it is possible to draw tentatively the following picture of the governing mechanism Estrin (estradiol) increases in the body fluids, due to stimulation of the ovary by the follicle-stimulating hormone of the pituitary, but when the concentration of this ovarian hormone reaches a certain level it acts in turn to suppress the output of the follicle-stimulating hormone, the concentration of estrogen upon which the proliferative stage of menstruation depends is thus reduced The hypophysis now releases its luteinizing principle which stimulates the development of the corpus luteum, but as the concentration of progesterone rises the production of the luteinizing principle is in turn suppressed, with the result that the integrity of the luteal tissue cannot be maintained The concentration of estrogen having fallen by this time to a low value, menstrual bleeding occurs and the

resumed—another cycle commences

THE FACTORS CONCERNED IN THE FERTILIZATION OF THE OVUM

The ovum if unfertilized degenerates within a few hours after its discharge from the ovary, probably in the Fallopian tube For this reason, and because only one ovum is discharged in each cycle, the period during which fertilization can occur is very short, probably not longer than 6 or 7 hours Yet the period during which coitus may lead to conception is relatively long, for the spermatozoa retain their power to fertilize the ovum for from 4 to 5 days If, for example, the sperm is deposited say on the 11th day of the cycle and ovulation occurs on the 15th, it is possible for the ovum to be fertilized The sperm shows motility for some time after it has lost its ability to fertilize The former property is therefore not a reliable index of the latter

As mentioned in the preceding section, the sperm must penetrate the ovum within a few hours after ovulation or cannot do so at all Fertilization is effected in the Fallopian (uterine) tube and here also the earlier stages in the maturation of the ovum take place When discharged, the ovum is surrounded by the *cumulus oophorus* It has been thought that this covering could not be penetrated by the sperm, and must first be destroyed by the enzyme hyaluronidase contained in the sperm of most mammalian species But there has been a misconception in this regard, for the male sex cell can reach the ovum through an intact cumulus It is possible, however, that the enzyme aids the sperm in some other way in its approach and penetration of the ovum An hyaluronidase inhibitor, such as tri-gentisic acid (rehibin) added to rabbit semen before insemination, was found by Parkes to prevent fertilization—an effect not due merely to a non-specific spermicidal action However, the administration of the hyaluronidase inhibitor to the animals was without effect upon fertility¹⁵

A CONSIDERATION OF THE FACTORS CONCERNED IN THE PRODUCTION OF THE MENSTRUAL FLOW

Hartman's studies have shown beyond doubt that menstruation occurs without ovulation in the

¹⁵ Martin and Bieler reported that the hyaluronidase inhibitor phosphorylated hesperidin administered to rats diminished fertility, Sieve later made the arresting announcement that this compound was anti-contraceptive in the human subject Chang and Pincus have been unable, however, to confirm these *in vivo* findings

monkey, especially between mating seasons. Menstruation without ovulation also occasionally occurs in women. A corpus luteum of course is not formed in these instances and the typical premenstrual endometrial changes do not occur. It appears, therefore, that, though the breakdown of a premenstrual endometrium built up under the influence of the luteal hormone ordinarily coincides with the menstrual flow and is normally part of the mechanism, it is not an essential feature. On the other hand, estrogen and the endometrial changes which it brings about are an indispensable part of the menstrual process.

Certain observations indicate that an important factor in the onset of menstruation is an abrupt fall in the estrogen concentration of the blood. (a) In monkeys, and also in the human subject ovariectomy (which removes the source of estrogen) is followed by uterine bleeding. This postoperative bleeding may be postponed by injections of estrin. (b) When a castrated monkey is given a series of injections of estrogen, uterine bleeding occurs a few days after the cessation of the treatment but unless the injections are long continued no bleeding occurs during the course of the treatment. The same phenomenon has been observed in women. Progesterone injections prevent the after effect of the estrogen treatment. (c) In women the maximum excretion of estrogen occurs in the intermenstrual period, and the minimum just before or at the time of menstruation. (d) Injections of a gonadotrophic extract of the anterior lobe of the pituitary into a monkey during the resting stage of the cycle causes bleeding a few days after the treatment has been discontinued. This result, since it cannot be produced after ovariectomy, is apparently due to estrin liberation and a rise followed by a fall in the estrin concentration of the blood. If the treatment with pituitary extract is followed by injections of estrogen, menstruation is postponed until a week or so after the latter treatment has been discontinued.

To summarize. Menstruation can occur in the absence of ovulation, the premenstrual endometrium characteristic of the action of the luteal hormone is therefore not necessary for its occurrence. Stimulation of the endometrium by estrogen however, appears to be essential, and a reduction in the concentration of this hormone in the blood seems to be the chief factor in the onset of the bleeding.

The role played by progesterone is by no means

clearly defined, but at least it can be said that the degeneration of the corpus luteum in ordinary ovular menstruation, and consequently, the cessation of progesterone production, coincides with or immediately precedes the menstrual flow.

ATTEMPTS TO IDENTIFY MENSTRUAL BLEEDING WITH ONE OR OTHER PHASE OF THE ESTRUS CYCLE OF LOWER ANIMALS

The menstrual cycle of primates is homologous with the estrus cycles of animals (expressed in terms of lower animals, the human is *polyestrus* (p. 876)), but to what stage of the estrus cycle menstrual bleeding itself corresponds is a debatable question. In animals such as the dog and cow bleeding occurs in pro-estrus. This led Heap to suppose that menstruation was

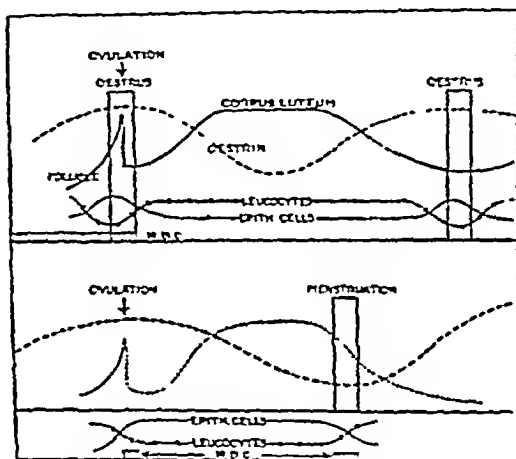


FIG. 61.12. Upper diagram, estrus cycle. Lower, menstrual cycle. (After Corner)

homologous with pro-estrus bleeding. But whereas the proliferative phase of the menstrual cycle corresponds apparently to pro-estrus in animals, menstruation itself does not occur until some 14 days after the termination of the proliferative phase (cf fig. 61.12).

It has been mentioned that the premenstrual changes in the uterus resemble and are considered to correspond to those of pseudo-pregnancy in animals. Parkes and Bellerby have therefore suggested that menstruation is the homologue of the slight bleeding which occurs in some animals as the uterine mucosa breaks down at the end of pseudopregnancy. Color has been given to this theory by the fact that uterine bleeding promptly ensues after destruction of the corpus luteum. It has just been stated, however, that menstruation can occur in the absence of a luteal phase.

Marshall has suggested that menstruation corresponds to the hemorrhage at the end of pseudopregnancy and the bleeding of the pro-estrus period of

the next cycle "telescoped" into one. The difficulty in correlating menstruation with pro-estrus bleeding has been mentioned. Furthermore, slight bleeding sometimes occurs in women at about the *time of ovulation* (intermenstrual period) which would be more comparable than menstruation with pro-estrus bleeding.

Another theory (*embryotrophic hypothesis*) postulates that menstruation is homologous with the extravasation of blood into the decidual tissues which occurs during implantation of the ovum and serves to nourish the embryo in the initial stages of its development. Occasionally the extravasated blood escapes into the uterine cavity and appears externally (the so called *placental sign*), a similar hemorrhagic discharge occurs in animals in a certain percentage of pseudopregnancies. This theory is in harmony with the well-known fact that in women bleeding may occur shortly after conception (i.e., during implantation of the ovum and at the time of the so-called "missed period") which is indistinguishable from that of an ordinary menstrual period. The theory is attractive in that it interprets menstruation as the culminating event in the preparation of the uterus for the reception of a fertilized ovum, the growth of the corpus luteum and the premenstrual changes in the uterine mucosa constituting the earlier preparatory processes projected in anticipation of implantation. See Corner, 1923, also Hartman, 1932.

It must be admitted that none of the foregoing views succeeds in harmonizing all the facts.

PUBERTY AND THE MENOPAUSE

The sexual life of the human female commences between the 12th and 16th years, when the first menstruation or *menarche* occurs. In hot climates the latter may appear as early as the 10th year. This period of sex awakening which is seen in animals and man of both sexes is referred to as *puberty*. The periods before and after are called *pre-* and *postpubertal*, respectively. The ovaries at this time show increased growth accompanied by maturation of the Graafian follicles and the discharge of ova. The growth of the uterus is accelerated, the breasts enlarge and, in both girls and boys, hair appears upon the pubes and in the axillae. Deepening of the voice, enlargement of the genitalia (penis and testes), increased muscular development and growth of hair on the face are characteristic features of this period in boys. In both sexes, somatic growth is accelerated.

Between the ages of 42 and 52 (average 47 years) the sexual processes come to an end. This time is spoken of as the menopause or climacteric. Menstruation ceases, retrogressive changes gradually supervene in the accessory organs of reproduction, e.g., atrophy of the uterus, shortening and

narrowing of the vagina, loss of epithelium and its replacement by fibrous tissue, and shrinkage of the mammary glands. These changes are the result of atrophic changes in the ovary—disappearance of the Graafian follicles together with a general fibrosis and shrinkage of the organ, similar changes in the uterus and vagina follow the removal of the ovaries in earlier life—*artificial menopause*. Psychic phenomena, usually mild, are not infrequent accompaniments of the menopausal period. Occasionally serious mental disturbances, e.g., melancholia, appear at this time. Vasomotor and other autonomic disturbances, hot flushes, sweating, etc., are very common features of the climacteric. The effect of the natural menopause upon the sex libido varies in different subjects. In married women it shows little alteration as a rule. Estrogen may be present in the blood and urine years after the onset of the menopause. The cause of the menopausal symptoms is not altogether clear, but there is little doubt that it is hormonal, gonadotrophic secretion is increased, whereas estrogen production is diminished. The follicle-stimulating hormone of the pituitary appears in the urine after the cessation of the menses.

MENSTRUAL IRREGULARITIES

The non-occurrence of the menstrual periods in postpubertal life is called *amenorrhea*, except during pregnancy, when suppression of the menses is a physiological phenomenon, the failure of the menstrual cycles at any time between puberty and the menopause is abnormal. Amenorrhea may be either *primary* or *secondary*. In the former instance the menses have never occurred, in the latter they appeared but were subsequently suppressed. Primary amenorrhea is in many instances associated with arrested development of the reproductive organs. Scanty menstruation is termed *oligomenorrhea*. *Dysmenorrhea* is painful menstruation. *Menorrhagia* is the term applied to excessive loss of menstrual blood. *Metrorrhagia* is the loss of blood from the uterus in the intermenstrual periods. A large proportion of these menstrual irregularities, when unaccompanied by some gross disease (tumor, etc.) in the uterus, have an endocrine basis. Amenorrhea is frequently the result of ovarian hypofunction. Dysmenorrhea is the result of irregular and spasmodic contractions of the uterine muscle which probably have also in many cases a hormonal origin. In the past, excessive bleeding in the absence of some obvious uterine disease has

been put down to inflammation of the uterine mucosa (endometritis). Before the important work of Hitschmann and Adler, who correlated the physiological changes in the endometrium with the stages of the ovarian cycle (p 890), the normal premenstrual characters of the uterine mucosa were considered to be pathological—evidence of endometritis¹

Shaw, from the study of a large series of irregular uterine bleedings unaccompanied by obvious disease of the uterus, came to the conclusion that only a very few could be attributed to inflammatory changes in the endometrium but that the great majority were due to ovarian abnormalities. In one group (*metropathia hemorrhagica*) the bleeding was continuous, the uterine mucosa was thickened and showed dilated glands and areas of necrosis, the constant finding in the ovary was an unruptured cystic follicle. The bleeding is apparently due to the overproduction of follicular hormone and the prevention of the formation of a corpus luteum. According to Kaufmann the imbalance between these hormones causes the proliferative phase of the endometrium (p 890) to become exaggerated, and prevents the onset of the secretory phase and the shedding of the mucosa as in normal menstruation. He describes the state of the endometrium as one of chronic cystic hyperplasia. This observer has produced a similar state of the endometrium in a castrated woman by the administration of large doses of follicular hormone.

In another group of cases of uterine bleeding reported by Shaw (*epimerorrhoea*) the menstrual flow was prolonged, that is, the intermenstrual periods were much shortened, the ovarian cycles, due to premature rupture of the follicles, were also abbreviated, ovulation occurring with abnormal frequency.

In cases of irregular uterine bleeding of endocrine origin the ovarian disorders are possibly due in turn to disturbances of a functional nature of the anterior lobe of the pituitary. It is well known that gross diseases of this part of the pituitary (e.g., acromegaly, Simmonds disease, dystrophia adiposo-genitalis, pituitary basophilism, etc.) are associated with menstrual abnormalities. These, however, usually take the form of oligomenorrhoea or amenorrhoea. Amenorrhoea also occurs as an accompaniment of hypothyroidism and as a symptom of several general diseases, notably anemia,

tuberculosis and mental conditions, malnutrition is not infrequently a cause.¹⁶

The *treatment* of menstrual disorders, supposedly of ovarian origin, by means of hormone preparations has been disappointing, and this is perhaps to be expected until a clearer understanding of the underlying factors governing normal menstruation has been arrived at. Estradiol dipropionate or benzoate has been employed with some measure of success in secondary amenorrhoea, especially if this has not been of long standing. Primary amenorrhoea is not permanently benefited by the follicular hormone. Werner and Collin produced uterine bleeding in castrated women by the administration of estrogen alone, but the premenstrual type of endometrium was not produced. Treatment of primary amenorrhoea with estrogen and progestin together would appear to have a more rational basis and might be expected to give better results, for, though the progestational development of the uterine mucosa is not essential for menstruation (p 892) this *does* occur in the normal subject. Kaufmann produced typical premenstrual endometrium in castrated women by the administration of the two ovarian hormones, but a total dose of over 1,000,000 international units of estrogenic principle and 35 rabbit units of corpus luteum hormone were required to produce this effect. He also induced growth of the uterus and uterine bleeding in cases of primary amenorrhoea associated with genital infantilism by the administration of large doses of estrogen over a period of months, or of this hormone followed by injections of the corpus luteum principle. A premenstrual endometrium was not produced in these cases and the bleeding was therefore not considered to be normal menstruation. In cases of primary amenorrhoea with well developed genitalia, a typical premenstrual endometrium followed by menstruation was induced by the administration of a total dose of from 1,000,000 to 1,500,000 international units of follicular hormone and from 35 to 50 rabbit units of corpus luteum hormone. In no case of primary amenorrhoea of the first group and in only one case of the second, however, was a cure brought about, in the sense that normally recurring menstrual cycles occurred, after the course of treatment, the subjects in all the other instances relapsed into their previous amenorrhoeic condition. Such a result

¹⁶ Vaginal smears, or more accurately, biopsies of the endometrium are employed in the diagnosis of menstrual disorders.

is to be expected, since the administration of ovarian hormones cannot correct the primary fault whether this be of ovarian or pituitary origin. In a number of cases of secondary amenorrhea, on the other hand, the spontaneous cyclic function of the uterus was restored by treatment with both hormones or with large doses of estrin alone.

Chorionic gonadotrophin of pregnancy urine has been used with some success where treatment with corpus luteum hormone (progesterone) is indicated, with the object of stimulating luteinization, and thus of increasing the patient's own supply of progesterone. But it does not appear to act in this way, there is no evidence that in the human this principle causes luteinization, it has not such an effect in the ape. Testosterone has also been used with success in excessive uterine bleeding.

Generally speaking, estrogens are employed to promote the menstrual flow when it is scanty or absent, and testosterone, or luteinizing hormones such as progesterone or chorionic gonadotrophin of pregnancy urine (HCG) either alone or in combination with estrogen or testosterone are employed when bleeding is excessive.

The greatest value of estrone or of the other estrogens is in the treatment of the nervous (psychic and vasomotor) disturbances of the menopause in which the estrogenic hormone concentration in the blood and urine is frequently definitely subnormal. Injections of estrogens in such instances relieve the vasomotor symptoms. Relatively small doses (200 international units daily) are required. Estrone has also been shown to be of definite value in the treatment of gonorrheal vulvovaginitis of children. Under its influence the vaginal epithelium undergoes cornification and, assuming the characteristics of the sexually mature woman, becomes more resistant to the disease. Estrin has also been employed with success in intractable cases of pruritis vulvae, and in certain ulcerative conditions of the vaginal mucosa.

We have seen (p. 882) that the changes in the uterine mucosa which are essential to the fixation and nourishment of the fertilized ovum are dependent upon the hormone of the corpus luteum. Deficiency of the latter may be responsible for certain cases of sterility or of habitual abortion, and when such a cause is suspected, treatment with an active preparation (which must always be given by injection) is a rational procedure. Its inhibitory effect upon uterine motility renders it a valuable agent in the treatment of dysmenorrhea, and its

action in suppressing the menstrual flow has been the basis for its use in menorrhagia and metorrhagia combined with estrogen. Excellent results have been reported following its use in the latter conditions.

The naso-genital relationship. A number of observations suggest a physiological relationship between the sex processes and the nose. In the first place the mucosa covering the conchae has a cavernous structure suggestive of the erectile tissue of the penis and clitoris, and olfactory stimuli and psychic aspects of sex are very closely associated. Nasal congestion, often accompanied by epistaxis, occurs regularly in many women at the time of the menses and in both sexes it is not unusual for nasal bleeding to occur at puberty. Sometimes the nasal hemorrhage in girls or women has seemed to replace menstruation which was coincidentally suppressed, and for this reason was termed erroneously "*vicarious menstruation*." Swelling and reddening of the nasal mucosa is a common finding in women during pregnancy and in monkeys during the estrus cycles. Stimulation of the interior of the nose has been reported to alter the periodicity of estrus cycles in rats, whereas excision of the conchae in young animals is said to result in hypoplasia of the sex organs. On the other hand, degenerative changes in the nasal mucosa have been observed as a sequence to castration, which could be reversed by estrogen injections. Finally, pseudopregnancy has been induced in rats by the nasal application of a strong solution of silver nitrate and by removal of the sphenopalatine ganglion. Mortimer, Wright and Collip became interested in the nasogenital relationship from the study of a French Canadian family, all of whom (both parents and nine children) suffered from *atrophic rhinitis*. An examination of the cranial skiagrams of these subjects disclosed signs which were interpreted as indicative of pituitary abnormalities. The possibility of a causative connection between the pituitary (e.g., deficiency of gonadotrophic principle) and the nasal disease suggested a trial of estrogen for the treatment of the latter. The local application of the follicular hormone to the conchae of these patients and of a number of other subjects of atrophic rhinitis is stated to result in definite improvement.

THE STRUCTURE AND DEVELOPMENT OF THE MAMMARY GLANDS AND THE SECRETION OF MILK

The Structure of the Mammary Glands

The secretory tissue of the mammary glands consists of elongated slender sacs or ducts—the *alveolar ducts*—whose walls show numerous pouches—the *alveoli*. The walls of the alveoli are composed of a basement membrane, a layer of myoepithelial cells, and a row of columnar epithelium in this order from without inwards. The epithelial cells are the secreting elements. The

gland is composed of some twenty-five lobes marked off by connective tissue septa containing much fat and derived from the mantle of adipose tissue which envelops it. A lobe is subdivided into lobules variable in size, shape and number. Each lobule gives rise to a narrow excretory duct which is the continuation of an alveolar duct but whose wall is quite devoid of alveoli. By the successive junctions of these lobular ducts a single *lactiferous duct* is formed which serves an entire lobe, it is lined by stratified epithelium. The lactiferous ducts converge toward the nipple. They become dilated beneath the areola and then constrict again. Near the base of the nipple they turn abruptly to run vertically toward the skin surface. Each opens separately at the apex of the nipple by a minute mouth called the *galactopore*. In the non-lactating gland few alveoli, if any, are to be found, the gland then consisting of ducts, connective tissue and fat. Preparatory to lactation the glandular tissue shows active growth leading to the development of the alveolo-lobular system which completely transforms the histological appearance.

The nipple and an area of skin surrounding it and known as the *areola* is darkly pigmented. The pigmentation deepens in the early months of pregnancy, a sign recognized from the earliest times. It only partially returns to its virgin color after gestation. As in other cutaneous pigmented areas the color is due mainly to the accumulation of granules of pigment in the squamous epithelial cells of the skin, but the actual production of pigment is a function of cells called *melanoblasts* lying just below the epithelial layers. These are the cells which give the "dopa" reaction (p 846), but they may themselves contain little pigment. The function of this pigment is unknown but its increase during pregnancy appears to be causally related to the functional activity of the underlying gland, an influence being conveyed from the active tissues to the melanoblasts through the rich sympathetic innervation (Cathcart and associates). There is possibly also a hormonal factor carried to the melanoblasts in the general blood stream.

The nipple contains smooth muscle which is arranged circularly as well as vertically. The vertical fibers are in close relation to the lactiferous ducts. The stiffening and erection of the nipple which results from mild stimulation is caused by the contraction of these strands of muscle. The nipple is one of the most richly innervated structures in the body. Nerve fibers both medullated (somatic) and non-medullated (sympathetic) are found in profusion in and beneath the skin, and especially in relation to the openings of the lactiferous ducts. In the dermis, nerve fibers form a loose open network without encapsulation, or enter end organs of various forms. Pacinian and Meissner's corpuscles commonly found in other cutaneous and subcutaneous areas are absent, and Krause's end-organs are scarce. The sensations aroused from the nipple and

areola are, apparently like those of the glans penis, of the diffuse protopathic or thalamic type. Light touch with cotton wool or van Frey's hairs is not appreciated.

The rôles played by the ovaries and the anterior pituitary in mammary development Prolactin (synonyms mammatropic, galactin, lactogenic hormone)

There are three phases in the development and activity of the mammary glands

(a) The mammary growth which is characteristic of puberty and of the commencement of subsequent estrus periods, is due, mainly at least, to the estrogenic hormone. In monkeys, treatment with estrogen will prevent the atrophy of the mammae which otherwise follows castration. This ovarian hormone brings the mammary glands of the virgin guinea pig, cow, goat, or monkey to full development, promoting the growth of both the alveolo-lobular and duct systems as well as that of the epithelium of the nipple.

Partial mammary development has been induced in male animals by the administration of estrogens. Petersen treated free-martins showing pronounced male characters with stilbestrol, and though in most of the treated animals there was no striking development of mammary tissue, in one, 21 lbs. of milk were secreted daily.

(b) The further enlargement of the mammary glands during pseudo-pregnancy and during the pregnant state is due to the growth of the corpus luteum. In most animals, although estrogen can stimulate growth of the ducts, stroma and nipple, development of the alveolo-lobular system can not, as in the case of the guinea pig, goat, cow, and monkey be induced by the administration of the follicular hormone alone. It must be combined with or followed by progestin administration, which then induces growth of the lobular-alveolar system. That is, the latter is superimposed upon the duct development induced by estrogen. The effect of the sex hormones upon mammary growth is enhanced by the administration of the thyroid hormone, which appears to be essential for the action of estrogen, stilbestrol is without effect upon the mammary tissue of thyroidectomized heifers.

An intact anterior lobe of the pituitary is essential for the stimulating action of the ovarian hormones upon mammary growth. In hypophysectomized animals, estrogen exerts little or no effect upon duct development unless treated with an anterior lobe extract. Full development of the

tubulo-alveolar system of hypophysectomized guinea pigs is induced by the daily implantation of the anterior pituitaries of rats which had been treated with estrogen. Stimulation of the duct and tubulo-alveolar system also results when immature ovariectomized rabbits are injected with anterior lobe material removed from cattle during pregnancy. Injections of pituitary material from non-pregnant virgin cattle is ineffective. Such observations led Turner and his associates to postulate the production by the pituitary of two mammaryogenic principles distinct from any known hormone, called *mammogen I*, which stimulated the duct system and was liberated under the influence of estrogen, and *mammogen II* which induced the growth of the tubular-alveolar system and whose production was controlled by progesterone. That is to say, the ovarian hormones, it was claimed, did not act directly upon the mammary tissue, but only through the medium of the pituitary principles. This theory has not received general acceptance, the evidence is far from being clear-cut, and the observed facts can be explained on the basis of direct control of mammary growth by ovarian hormones in conjunction with the action of known anterior lobe hormones, prolactin, ACTH and the growth hormone. The first of these appears to exert, in combination with the ovarian principles, an essential effect, while the other two serve in a non-specific way to maintain the health and vigor of the animal. The specific action of prolactin in promoting mammary growth (as distinct from its lactogenic effect, described below) was shown by Lyons who induced localized hyperplasia of the glandular epithelium by injections of prolactin into the mammary ducts of rabbits previously treated with ovarian hormones.

The prolactin effect upon mammary development is a dual one, the direct effect is aided by an indirect one, namely stimulation of progesterone production (luteotrophic action).

The nature of the direct prolactin effect is an unsettled question, it may act to sensitize the gland to the ovarian hormones, or it may act synergically with them.

Androgens, such as testosterone, stimulate mammary growth in both male and female animals, an effect which requires an intact hypophysis.

The role of the placenta in mammary growth

The several factors which have been mentioned are not able to induce a degree of mammary growth

equal to that occurring naturally during pregnancy—a fact which at once suggests the participation of the placenta in the process. Removal of the ovaries and fetuses from rats together with hypophysectomy causes mammary regression if the placentae are also removed, but mammary growth continues if the placentae are retained. Also, mammary involution occurs if the placentae are removed even though the ovaries and pituitary are left intact. It seems then that the placenta is indispensable for full mammary development and can substitute for the ovaries as well as for the pituitary.

(c) The actual *secretion* of milk which occurs at the end of pregnancy is brought about through the pituitary. A lactogenic effect of anterior lobe extracts was demonstrated by Grueter and Stricker in 1929, and Corner in 1930 showed that injections of anterior lobe extract caused hypertrophy of mammary tissue and secretion of milk in ovariectomized virgin rabbits. Riddell and his associates demonstrated that the lactogenic effect is due to a separate hormone of the anterior hypophysis. They obtained extracts of the anterior lobe which had pronounced lactogenic effects but were practically free from thyrotrophic, growth and gonadotrophic principles. They named the hormone *prolactin*. It has since been obtained in crystalline form, and in the form of what appears, from electrophoretic and diffusion studies, to be a pure protein. It is most probably a product of the acidophil cells.

Prolactin induces proliferation of the epithelium lining the crop glands of male and female pigeons and doves,¹⁷ and increases the production of a caseous material, consisting of desquamated epithelial cells and known as "crop milk", with which the young are fed. The epithelium becomes heaped up into a number of layers, producing pronounced thickening of the mucosa, an effect which is strikingly evident even upon gross inspection. Great numbers of mitotic figures appear. The stimulant action upon epithelial hyperplasia is greatly magnified by the administration of colchicine, a drug which arrests mitosis in the metaphase stage. The maximum effect of the drug is manifested in from 6 to 8 hours and persists for a period of from 16 to 18 hours. The lactogenic hormone

¹⁷ This action is employed as a means of assaying the potency of lactogenic extracts. In general terms a "bird unit" is the minimal quantity of extract required to induce a certain increase in weight of the crop glands of doves or pigeons.

causes the mammary secretion in all species of mammals investigated. On the other hand, hypophysectomy suppresses milk secretion, and in the cat, though pregnancy is not terminated by hypophysectomy performed in the later months, lactation does not occur postpartum in the absence of the pituitary. Prolactin arouses the maternal instinct in virgin rats, Riddle and his associates have shown that, in from 5 to 12 days after a series of injections, the treated animals will care for young. It also induces broodiness in hens. Continued injections suppress the gonadotrophic action of the hypophysis with consequent atrophy of the gonads. In pigeons prolactin manifests a growth effect upon the body as a whole, but especially upon the liver and intestines and probably upon the pancreas (splanchnomegaly).

That milk secretion in the human subject is also under the control of the anterior lobe is indicated by the fact that in acromegaly (p 799) milk secretion may persist for an extraordinarily long time after childbirth (5 years in a case of Cushing's). It may also occur in male subjects of gigantism (p 799). These observations are most readily explained upon the assumption that in these pituitary abnormalities, the overactivity of the anterior lobe causes, as well as an overproduction of growth hormone, a continuous secretion of prolactin, corticotrophin and possibly other lactogenic principles. Prolactin preparations have been used clinically to increase the milk flow. Riddle and his associates report that in a series of twenty-nine parturient women the daily milk secretion was increased in twenty-five of them by from 50 to 400 grams in from three to nine days after treatment was instituted.

Prolactin has also been found in the placenta, in the urine of lactating women, and in the urine of babies secreting "Witch's milk" (see below).

The production of the lactogenic hormone by the pituitary is stimulated by estrogen, and its liberation from the gland is brought about, apparently reflexly, by the act of suckling.

A CONSIDERATION OF FACTORS CONCERNED IN THE INITIATION OF MILK SECRETION AT THE TERMINATION OF PREGNANCY

Though, as mentioned above, in most animals as well as in the human subject, mammary growth occurs during pregnancy, the flow of milk is not established until after the birth of the young and the expulsion of the placenta. Complete information concerning the factors which initiate

milk secretion after the birth of the child is not available, though the importance of prolactin in the secretion of milk has been proved, the influences determining its discharge from the pituitary and the part played by other hormones are, to a large extent, unknown.

The *hormonal* and *nervous factors* which may be concerned in the induction of mammary secretion postpartum will now be reviewed.

HORMONAL. As an outcome of a number of observations the *hormonal release theory* of mammary secretion was proposed. (1) When mammary growth is induced in the guinea-pig by the injection of estrogen, secretion of milk occurs when the dosage of hormone is suddenly reduced. (2) Collip and his colleagues reported that if in virgin rats the luteinized ovaries induced by the administration of human chorionic gonadotrophin were removed, the hypertrophied mammae secreted profusely. Removal of the pituitary, however, together with the ovaries prevented this result. (3) Lactation is inhibited in most animals by estrogen, or by measures which stimulate estrogen liberation, it is also promptly suppressed by HCG which stimulates the production of luteal tissue. (4) When lactating cows become pregnant, milk production declines progressively throughout gestation but rises again abruptly after calving. (5) It is said that reduction in the estrogen concentration of the blood of parturient women, as may be induced by the administration of a diuretic such as theophyllin, stimulates milk secretion. These are among the observations which have suggested that in the intact animal, lactation is held in abeyance by a high concentration in the blood of estrogen, and possibly progesterone and a placental principle, but that with the fall in concentration of these hormones toward the end of pregnancy, or postpartum, the pituitary and the adrenal cortex (see below), now no longer restrained, exert their lactogenic action. Evans also explains the milk secretion which occasionally occurs in the child shortly after birth (witch's milk) on this basis. The blood of the fetus presumably contains ovarian hormones of the mother and the elimination of these after birth results in the liberation of prolactin from the infant's pituitary. The theory just outlined concerning the induction of lactation fails to explain the phenomenon in all species. In some forms (goats and cattle) continued treatment with estrogen in small doses actually stimulates mammary secretion.

Reduced prolactin content of the pituitary or of

the blood under the influence of estrogen injections, which would be consistent with the theory, has not been observed (Meites and Turner)

There are other experimental findings which argue against the theory based upon the release of the pituitary from hormonal inhibition. Two alternative theories have been proposed. The first is that of Petersen, and probably has the greater amount of experimental support. It is now generally accepted that milk production, i.e., the formation of milk and its secretion into the alveoli begins some time before parturition. The question, therefore, is not a matter of *secretion* after childbirth, but of the discharge or *ejection* of milk from the gland. That the prepartum secretion is considerable is testified to by the swelling of the breasts in the later part of pregnancy. Also, as pointed out by Petersen, the low water content of the first milk secreted, the *colostrum*, strongly suggests a retained secretion. According to Petersen, the first milk is ejected as a result of stimulation of smooth muscle in the vicinity of the alveoli and ducts, or of the myoepithelial cells in the alveolar walls, by posterior pituitary principles, especially by pitocin (oxytocin) but possibly also by pitressin. As shown many years ago by Ott and Scott, the oxytocic principle has a powerful galactagogue action. Pitressin is much less active in this respect. The stimulus for the release of pitocin is provided presumably by afferent nerve impulses set up in the nipple by the act of suckling, a conclusion which would account for the very rich innervation of the nipple. This theory, then, implies that during the later part of pregnancy, the retained secretion tends by distension of the alveoli to depress secretory activity, but with the emptying of the alveoli, the inhibitory effect of distension is removed, and a continuous profuse secretion is maintained by the lactogenic hormone.

The experiments of Petersen and his associates indicate clearly that a nervous-hormonal mechanism is responsible for the phenomenon known as the "let down" of milk in cows. The blood from animals who had "let down" their milk when perfused through excised mammary glands caused a profuse secretion of milk within 15 seconds. Blood from cows which had not "let down" their milk was quite ineffective. Fright, nervousness, or the injection of adrenaline prevented the "let down" phenomenon. The well-known stimulating action of suckling upon the uterine contractions is also strongly suggestive of the liberation of pitocin into the circulation.

According to the second alternative theory,

which has been proposed by Meites and Turner, the chief factor in the initiation of secretion postpartum is the great increase in prolactin production which occurs at this time. In rats and guinea pigs the prolactin content of the pituitary was found to increase some four-fold a few days before the birth of the young, and in women the urinary excretion of lactogenic hormone postpartum is some 16 times greater than during pregnancy. Estrogen was found to increase the production of prolactin by the hypophysis, the latter's low prolactin content during pregnancy is attributed to the relatively high progesterin content of the blood and its antagonistic action upon the activity of estradiol. It is postulated that the stimulating effect upon the pituitary is dependent upon the estrogen-progesterin ratio in the circulation. At the end of pregnancy a sharp reduction in progesterin concentration occurs. It was found that estrin and progesterin administered in appropriate amounts had no effect upon the quantity of prolactin in the pituitary.

In milking cows, thyroidectomy causes a reduction in the quantity of milk and, as a consequence, a diminution in the total yield of milk fat (Graham). The administration of the thyroid hormone, of thyrotrophin or of iodinated casein increases the quantity as well as the fat content of the milk. These effects appear to be specific and not due simply to changes in metabolism, for raising the heat production 50 per cent or so above normal by means of dinitrophenol decreases both the volume and the butter fat of the milk. The effect of thyroidectomy is attributed, in part, to the removal of the parathyroids.

The adrenal cortex plays an essential part in milk secretion (see Hartman and associates). Rats, adrenalectomized in the later months of gestation fail to produce the normal quantity of milk. The defect can be corrected by the administration of very large doses of adrenal cortical extract, but not by ordinary maintenance doses, and also, not infrequently, by a high salt diet. The effect of adrenalectomy upon mammary secretion may be to a certain extent non-specific and result from the disturbance in mineral, water and carbohydrate metabolism, the drying up of milk production being an expression of the dehydrated state. Nevertheless, that the adrenal cortex plays a specific rôle in milk-production appears likely from the fact that prolactin alone has no lactogenic action in hypophysectomized animals unless an extract of the adrenal cortex (or corticotrophin) is administered with it. This leads to the belief that normally

the pituitary brings about mammary secretion by liberating ACTH as well as prolactin. Adrenal cortical tumors in males are sometimes associated with gynecomastia—further suggestive evidence of the role of corticoids in lactation. Of the adrenal principles, desoxycorticosterone has but slight lactogenic action, whereas, 11-dehydro-17-hydroxycorticosterone is effective. The existence of a specific lactogenic hormone, called *cortilactin*, has been postulated, but not proved. The importance of the adrenal cortex in lactation and certain recent observations with the glycotropic factor of the pituitary, as well as the influence of the ovarian hormones have led many, especially Folly and Young, to doubt the specificity of prolactin itself. The hormonal control of lactation is probably more complex than has been thought and involves perhaps several hormones.

NERVOUS Cannon and his associates have shown that sympathectomized animals (cats) are in most instances unable to suckle their young (ch 72), a small quantity of milk is secreted after the young are born but the secretion soon dries up, maternal instinct is abolished. It is also well known that after weaning the secretion of milk ceases, an event which may be due to (a) removal of the stimulus to the nipples provided by the act of suckling, or (b) distension of the alveoli of the gland by accumulated milk. Selye found that it is the first rather than the second of these factors which is of greater importance. The main duct of the gland was ligated in rats but the young were allowed to suckle. The accumulation of milk caused marked distension of the alveoli but secretion was not inhibited. On the other hand, removal of the sucklings from the mother or excision of the nipples suppressed secretion. It is probable that the effect upon mammary secretion of stimulating the nipple is not a local one but is the result of afferent impulses reaching the pituitary and causing the liberation of prolactin. Such a mechanism would be analogous to that by which in certain animals a pituitary influence called into play by stimulation of the genital tract, induces ovulation (p 886).

Another possible influence was suggested some years ago by Selye and his associates, namely, that inhibitory afferent impulses from the distended uterus held lactation in abeyance. They reported that, in the rat, emptying the uterus before the end of the term caused milk to be secreted but that, if after its evacuation, the uterus were maintained in the distended state by

melted paraffin wax (with melting point a little above body temperature), lactation did not occur. However, the inhibition observed should probably be attributed to the effect of the operation upon the well being of the animals and therefore is not specific. Bradbury and Greene, who repeated these experiments, found that secretion occurred after the animals had recovered from the operation though the uterus was still filled with wax.

The experiments cited above indicate the importance of *afferent* nervous influences in milk secretion, presumably through the liberation of prolactin. On the other hand, the fact that secretion may continue after complete denervation of the glands and that mammary tissue transplanted into a pregnant animal secretes milk after the young are born provides clear evidence that neither *afferent* or *efferent* nerves are essential to the secretory mechanism. Nevertheless, continued activity of the gland and the maintenance of a normal milk flow appear to be dependent upon prolactin and possibly other hormones kept at a high level of production by stimulation of the nipple. The principles of the posterior lobe, though perhaps not essential, appear certainly to facilitate the initiation of lactation.

THE COMPOSITION OF MILK, THE ORIGINS AND SECRETION OF ITS CONSTITUENTS

With the onset of lactation the epithelial cells lining the alveoli become loaded with fine droplets of fat which soon coalesce to form large globules in the half of the cell lying next to the alveolar lumen. The fat with the other constituents of the milk is then discharged into the alveolus by constriction of the cell on the basal side of the globule. Thus, the inner, free part of the cell is pinched off as it were from its nuclear or basal portion (*apocrine* type of secretion). In other instances the inner boundary of the cell appears to rupture and discharge its secretion. The blood supply to the gland during lactation is profuse, amounting in the cow, according to Graham and his associates, to 400 volumes of blood to 1 volume of milk produced.

The quantity of protein in cow's milk is more than double that in human milk, but the sugar and fat contents are lower. See table 85. The proteins of human milk are, however, of higher biological value. In cow's milk casein constitutes from 80 to 90 per cent of the total protein, and lactalbumin from 10 to 20 per cent, whereas in human milk from 30 to 40 per cent of the total protein is lactalbumin. According to Murlin, all the protein

of mother's milk may be retained by the infant, a considerable proportion of the protein of cow's milk, on the other hand, is eliminated. Of the essential minerals, calcium is present in largest amounts, namely, about 0.4 per cent.

The proteins of milk are not found in the blood, yet only a small fraction of the former can be accounted for by synthesis from the blood amino-acids. The milk proteins may possibly be derived from the plasma proteins through some chemical transformation brought about by the activity of the glandular tissue. A more likely origin is from a *glycoprotein* present in small amounts in blood. It has been shown by Turner and his colleagues that this protein disappears from arterial blood supplying the glands. The carbohydrate part of the molecule, which consists of galactose, mannose and glucosamine, and constitutes 9 per cent, may be used for the production of lactose, while the remainder might serve as a precursor of the milk proteins. The phosphorus required for the synthesis of casein (a phosphoprotein) is derived from the inorganic phosphorus of the blood.

Milk fat differs in composition from the neutral fat of the blood in having a relatively high proportion of short-chain fatty acids of the series C_4 to C_{14} . The fatty acids, e.g., butyric, caproic, caprylic, lauric and myristic, constitute about 30 per cent of the total fat of milk. The fat of milk therefore cannot be simply neutral fat transferred from the circulation to the gland and there concentrated. Of the long chain fatty acids, oleic (35 per cent) and palmitic (30 per cent) are present in greatest amounts. Stearic acid constitutes only about 2 per cent. The origin of the short-chain fatty acids is unknown. It has been shown that in ruminants, short-chain fatty acids are formed in the rumen through bacterial action upon carbohydrates, and that such acids are absorbed into the blood stream, the suggestion has, therefore, been made that they are concentrated by the mammary gland. But an objection to this possibility is mentioned by Kay, namely, the observation of Malpress, that sodium butyrate fed to lactating cows does not increase the proportion of volatile fatty acids in the milk.

Of the source of the main solids of milk most is known definitely about that of lactose. It is generally accepted that it is derived mainly from the glucose of the blood. From the evidence no other conclusion can be reached. No significant amount of glucose is present in milk, and the arterial blood,

in its passage through the lactating gland loses over 25 per cent of its glucose content, also, surviving slices of mammary tissue, incubated with a solution of glucose, synthesize lactose, and, any lowering of the blood glucose, as by the administration of insulin, is reflected in the lactose content of the milk. Mammary tissue contains about 0.2 per cent of glycogen, which is increased under the influence of insulin, and it may well be that it plays an intermediate rôle in the production of the milk sugar.

Both cow's milk and human milk vary greatly both in composition and quantity. Many factors, psychic, dietary, hormonal, period of lactation,

TABLE 85

	TOTAL PROTEIN	CASEIN	LACTAL- BUMIN	LACTO- GLOBULIN	LACTOSE	FAT	ASH	WATER
	per cent	per cent	per cent	per cent	per cent	per cent	per cent	per cent
Cow's milk	3.6	3.0	0.5	0.08	5	3.5	0.7	87.2
Human milk	1.5	1.1	0.4	—	6.8	4.0	0.2	87.5

time of day (greater volume during the night), etc., exert their effects upon the quality and quantity of human milk, the daily output increases up to a maximum of from 1000 to 3000 cc. at about the 25th week after childbirth and then gradually declines. Human milk of good quality contains all the elements essential for building body tissue, and adequate amounts of vitamins A (as well as carotene), the B complex and C. Vitamins B₁ and D are in low concentration and may be present in inadequate amounts. The average composition of cow's milk and woman's milk of good quality are compared in table 85. The *quality* and *quantity* of the milk is influenced by diet. An ample diet tends to maintain both at a high level. Raising the protein in the diet tends to increase the total milk yield. A high fat diet increases the yield as well as the fat content of the milk. Carbohydrates, if in high proportion in the diet reduce both the yield and the quality of the milk.

THE TESTES

Structure of the testes

The substance of the testes consists of a mass of coiled tubules—the *convoluted seminiferous tubules*—

bound together by a *stroma* of connective tissue. The seminiferous tubules composing the body of the testis consist of a number of closely coiled tubes from 35 to 70 centumeters long and from 150 to 300 micra in diameter. They are held in wedge shaped compartments or *lobules* by incomplete connective tissue septa. Through the gaps in the septa they are joined to one another by short tubular connections. The combined length of the convoluted tubules of man has been estimated at over 300 yards. At the apices of the lobules the seminiferous tubules join the first elements of the excretory duct system. These are narrow straight tubules—the *tubuli recti*—which empty into a system of irregular cavernous spaces—the *rete testis*—in the *hilum* or *mediastinum* in the posterior part of the testis. From the rete testis a number (8 to 15) ducts, called the *ductuli efferentia*, arise, which passing to the head of the epididymis become confluent to form the *ductus epididymis*. The latter is a long, greatly tortuous channel which forms the body and tail of the epididymis. At the lower extremity of the epididymis the ductus bends abruptly upon itself to be continued as the *vas* or *ductus deferens*. The vas deferens on each side, after a devious course, is joined by the duct of the corresponding seminal vesicle, and terminating as an *ejaculatory duct* opens into the urethra. The *sperm* or semen is stored mainly in the ductuli efferentia and in the epididymis.

The convoluted tubules are lined by several layers of cells. With the exception of a relatively small number of supporting cells—*cells of Sertoli*—the lining cells develop through several distinct stages into spermatozoa. They are, therefore, called the *spermatogenic* or *germ cells*. In the mature testis the several stages in the development of these cells can be distinguished. They are arranged in layers from without inwards (i.e., toward the lumen of the tubule) in an order, according with the degree of their maturity. The most primitive of the germ cells are large ($12\ \mu$ in diameter) round cells lying upon the basement membrane. They are called *spermatogonia*, they alone are present in the immature testis. The first cells to be developed as a result of the activity of the pituitary are called *primary spermatocytes*. These are larger than the spermatogonia (15–19 micra), each divides into two smaller *secondary spermatocytes*, each of which divides in turn to produce two *spermatids*. Each spermatid contains only half the number (haploid) of chromosomes in its nucleus. This last step in maturation is called *reduction division*, or *meiosis*, but the reader is referred to texts on genetics for a description of this process and for further details in spermatogenesis. The spermatids which are found near the lumen of the tubule are transformed directly, i.e., without division, into the free *spermatozoa* (see fig. 61.13). About half the number of spermatids (or spermatozoa) contain a Y-chromosome, and half contain an X-chromosome.

The semen contains some 60,000,000 spermatozoa per cubic centimeter, which are suspended in a fluid the

great bulk of which is secreted by structures outside the testes, ductuli efferentia, ductus epididymis, seminal vesicles, prostate, and urethral glands. The spermatozoa in the seminiferous tubules are non motile. They are carried along the ductuli efferentia by the cilia of the cells lining these channels. Reaching the ductus epididymis they may here remain for a long time, or until an ejaculation occurs. Ejaculation of the semen during coitus is brought about by contractions of the vasa deferentia and the ejaculatory ducts. In the stroma of the testis, isolated groups of polyhedral epithelial-like cells are present—the *interstitial cells* or *cells of*

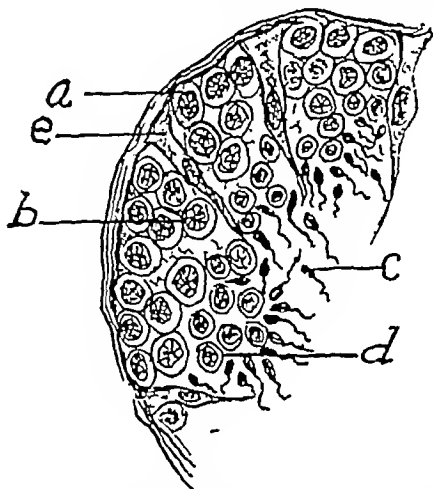


FIG. 61.13 Spermatogenesis, a, spermatogonia, b, primary spermatocytes, c, spermatozoa, d, spermatids, e, cells of Sertoli

Leydig. These contain lipoid granules in their cytoplasm, and are stimulated, apparently by the luteinizing hormone of the anterior pituitary.

THE PHYSIOLOGY OF THE TESTIS—THE MALE HORMONE

The testis has two functions, namely, the production of germ cells—*spermatogenic function*—and the elaboration of male hormone—*endocrine function*.

The results of castration and transplantation experiments, an account of which has been given at the beginning of this chapter (p. 871) leave no doubt that the testis is a gland of internal secretion, though the earlier attempts to extract the testicular hormone were unsuccessful. The chief cause of failure was probably that a reliable test object had not been devised. John Hunter, as long ago as 1792, showed that the normal function of the prostate and seminal vesicles was maintained only in the presence of the testes, and Berthold, in 1849,

demonstrated that the testes discharged some material into the blood stream essential for the growth of the cock's comb (see also the beginning of this chapter) Among the first to attempt the preparation of an active material was the noted physiologist Brown-Séquard (1889) who administered testicular extracts to himself and thought that he acquired an increase of vigor and a greater capacity for work after the treatment The idea that senescence was related to testicular atrophy and the consequent reduction or loss of the male hormone was revived a few years ago by Voronoff in France and Steinach in Germany Transplantation of apes' testes into old men have been performed and claims have been made upon the most flimsy evidence that the grafted tissue lives and causes rejuvenation of the recipient There is no reason to suppose that testicular atrophy has any causal relationship to old age phenomena Indeed, it is well known that in the male the degenerative changes of the body incident to age usually far outstrip the decline in the sexual functions In the female, on the other hand, regressive changes in the ovaries occur normally after a certain age, yet this event does not appear to hasten senile changes Nor do eunuchs show premature senility Furthermore, the bodily changes associated with declining years are degenerative and irreversible, it is therefore irrational to hope that rejuvenation can be brought about through grafted testicular tissue Moreover, any hormone effect which might result from such a procedure would be only temporary, for the graft does not live (p 873)

The spermatogenic function of the testis is affected adversely by several conditions, namely, repeated or prolonged exposure to X-rays or to radium emanations, alcoholism, vitamin E or B₁ deficiency, close confinement (in certain animals) and an elevation in temperature of the testis tissue amounting to only a few degrees In the developmental anomaly known as cryptorchidism the testis fails to descend into the scrotum, spermatogenesis does not occur in the undescended organ The defect is due undoubtedly to the relatively high temperature of the abdominal cavity which is considerably higher than that of the scrotum Raising the testis of the normal adult animal from the scrotum and fixing it in the abdomen is followed within a few days by degeneration of the seminiferous tubules But the latter are restored to their normal appearance and function if the testis is returned to the scrotum, provided that its stay in the abdomen has not been too long

It has also been found that exposure for 15 minutes or so to a temperature of 6°C above the normal body temperature leads within 10 days or so to degeneration of the sperm-producing cells The latter effect follows in the ram if the scrotum is enclosed in heat-insulating material, such as coverings of wool, so as to maintain the scrotal temperature a few degrees above the normal level Though the human testis may suffer in febrile diseases as a result of the high temperature the loss of function is, as a rule, only temporary

The internal secreting function of the testis is resistant to the conditions mentioned above The undescended testis continues to secrete its hormone, for castration effects are not seen in bilateral cryptorchidism, nor does exposure to the X-rays or vitamin deficiency lead to a failure in the endocrine function of the gonads

Ligation of the vas deferens has been stated by some (e.g., Steinach) to cause degeneration of the spermatogenic cells and an increase in hormone production These claims have been conclusively disproved Vasoligation, which has been advocated as a means of increasing the output of male hormone, is based upon error and is never justified for this purpose, it exerts no detectable effect either upon the sperm-producing cells or upon those of internal secretion

✓ In many wild species the activity of the male reproductive organs is confined to a definite mating season The changes occurring during this period of "rut", as it is called, are analogous to those which take place during the estrus periods of the female The testes hypertrophy, spermatogenesis occurs, and sexual desire is aroused The anterior pituitary exerts a governing influence upon the testes essentially similar to that which it exerts upon the female gonads (p 886)

Pezard as early as 1911 produced comb growth in capons by the injection of a saline extract of testicular tissue McGee in 1927 obtained an active lipoid extract of bull's testes As a result of the subsequent studies of this worker and his associates and those of other investigators (Moore, Gallagher, Koch) the following effects of testicular extracts have been demonstrated

(1) Growth of the comb, wattles and ear-lobes of the castrated cockerel (capon) and of the comb in hens (fig 61 14)

(2) Inhibition of ovulation in hens

(3) Prevention of the atrophy of the accessory male organs (seminal vesicles, prostate, Cowper's glands) in the castrated guinea pig, rat or mouse

(fig 6115), the maintenance of the pendulous form of the scrotum, and restoration of the electrical ejaculatory response in the hypophysectomized guinea pig.¹⁸ Testosterone mainly stimulates the glandular elements of the prostate

when formed are non-motile, becoming so only upon reaching the epididymis where, apparently, they are acted upon by the male hormone. Spermatogenesis in the immature (33 days old) rat can also be induced by injections of testosterone



FIG 6114 Showing the effect of testis hormone (from urine) upon the comb growth of capons. The birds in the upper photograph received daily injections over a period of 15 days. Lower photograph, untreated controls (After Funk, Harrow and Lejwa)

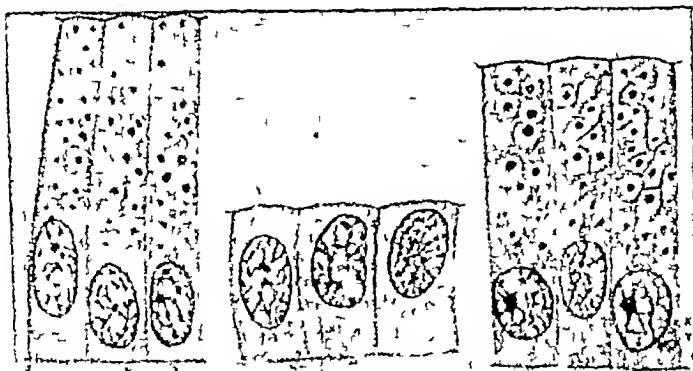


FIG 6115 Effect of castration and testis hormone on epithelium of seminal vesicles (Moore, Hughes and Gallagher) 1, cells from normal animal, showing secretion granules, 2, cells from twenty day castrate, 3, cells from twenty day castrate treated with male hormone.

(4) Extension of the lives (as judged by motility) of the spermatozoa in the epididymis of hypophysectomized guinea pigs. This fact suggests that the testicular hormone normally performs an important function with respect to the life and motility of the male sex cells. The spermatozoa

¹⁸ Passing an alternating electric current (30 volts) through the head of a normal anesthetized guinea pig causes an ejaculatory reflex and the discharge of secretions from the vas deferens, seminal vesicles and prostate. This response was discovered by Batelli in 1922, it is abolished after castration.

(5) Reduction in the size of the testes and atrophy of the seminiferous tubules by repeated injections. This effect is probably exerted through the hypophysis.

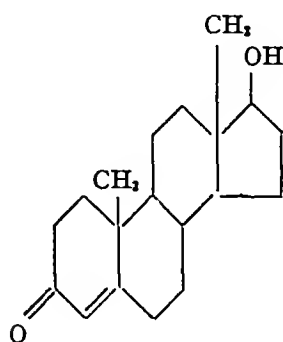
(6) Restoration to normal of the "castration cells" of the anterior pituitary (p 887)

(7) Masculinization of female rat fetuses following repeated injections during pregnancy, hypertrophy of the uterus and reduction in the size of the ovaries of newly-born rats.

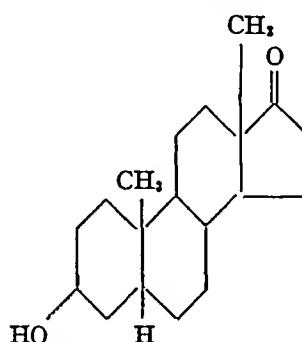
The testis hormone belongs to the class of

sterols, being closely related chemically to the female sex hormones, to cholesterol, the bile salts, etc (see p 878) It has been obtained in crystalline form from testicular tissue, this product, which is regarded as the true male hormone, has been named *testosterone*. It has been synthesized from cholesterol. Butenandt isolated a crystalline androgenic compound from urine. This is much less active (from $\frac{1}{7}$ to $\frac{1}{10}$) than testosterone and differs from the latter in chemical structure as shown in the formulae given below, it is called *androsterone*. Both testosterone and androsterone have been prepared from cholesterol by Ruzicka and his colleagues

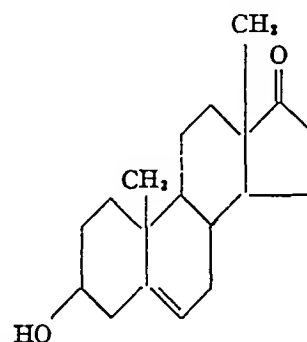
(probably estrone) is excreted by the male¹⁹ The androgen/estrogen ratio (i.e., the number of international androgenic units excreted per day over the estrogenic activity in micrograms of excreted estrone per day) is from two to five times higher in males than in females (see Gallagher and colleagues). The main source of the androgens in female urine is the ovary, for ovaries grafted into the ears of castrated male rats maintain the seminal vesicles and prostate in a normal state (Hill), yet excretion is not entirely abolished by ovariectomy. It must be borne in mind that the male and female hormones are closely similar in chemical structure, estradiol can be converted to testosterone by



Testosterone ($C_{19}H_{28}O_2$)
(3-Keto-17-hydroxy- Δ^4 -androstene)



Androsterone ($C_{19}H_{30}O_2$)
(3 Cis-hydroxy-17-keto-androstane)



Dehydroandrosterone ($C_{19}H_{28}O_2$)
(3-Trans hydroxy-17-keto- Δ^4 -androstene)

A second androgenic compound, *dehydroandrosterone*, is present in urine, its potency is about half that of androsterone. These urinary androgens are metabolic products of testosterone. The latter is also degraded to etiocholanol-3(α)-one-17, which appears in the urine after the administration of the testis hormone. Acid hydrolysis of the latter in the process of extraction liberates the active principles *Androstenedione*, $C_{19}H_{26}O_2$, and *androstenediol*, $C_{19}H_{30}O_2$, are laboratory products, the former having been prepared from androsterone, the latter from dehydroandrosterone. Esterification of testosterone, as demonstrated by Parkes, greatly increases and prolongs its action, testosterone acetate and propionate, especially the latter, being several times more effective than the free hormone.

Testosterone is most effective when administered by injection, it is also effective percutaneously (byunction) and by mouth but several times the injected dose is required. Androgenic substances (androgens) first make their appearance in male urine at the time of puberty or somewhat earlier. Androgens, surprisingly enough, are also present in female urine in amounts not greatly less (4) than in male urine and, as already mentioned, estrogen

dehydrogenation, and testosterone injected into a male is excreted partly as androgen and partly as estrogen. Furthermore, progesterone in large doses is capable of androgenic effects (the maintenance in castrates of spermatogenesis and enlargement of the prostate). The adrenal cortex furnishes a proportion of the androgens and estrogens found in normal urine, and all of those excreted after castration or ovariectomy. The urinary excretion of androgens is reduced to about half the normal in eunuchoidism, and to even less than this after castration. In adrenal virilism the excretion of androgens (17-ketosteroids) is considerably increased (ch 59).

Several methods have been proposed for assaying the potency of testicular preparations. Gallagher and Koch use the comb-growth test. They define a unit of testes hormone as the amount of extract which when injected daily for 5 days yields an average of 5 mm increase in size (length plus height) of the combs of at least 50 per cent of a group of brown leghorn capons.

¹⁹ In some instances a part of the androgenic activity of urine, especially in cases of virilism (p 839) may be due to the presence of *adrenosterone*, an androgen found in the adrenal cortex.

The interstitial cells of Leydig

In animals which show a rutting season the male hormone is secreted intermittently, in others, e.g., the rat and guinea pig, and in man, the secretion is continuous. The elements of the testis which elaborate the hormone are not known with certainty, though it is generally believed that the interstitial cells of Leydig serve this function. The following is a summary of the evidence usually cited in support of this view.

(a) Under any of the following circumstances the seminiferous epithelium undergoes atrophy, but the interstitial cells remain intact, and castration effects do not appear: (i) treatment of the testes with mild doses of X-rays, or ligation of the vas deferens, but leaving the spermatic vessels intact, (ii) excision of the testes and grafting them elsewhere in the body, (iii) non-descent of the testes into the scrotum (cryptorchidism), the degenerative changes in the spermatogenic cells are probably due to the high temperature to which they are exposed in the abdomen.

(b) Conversely, damage to the interstitial cells without injury to the seminiferous epithelium can be induced in rats by the administration of pitch, or by a diet lacking in the vitamin B complex, atrophy of the prostate and seminal vesicles results.

Hypogonadism and hypergonadism in the male

Tumors of the testis in young boys may result in the precocious development of the secondary sex characters—growth of hair upon the pubis, in the axillae and over the face, deepening of the voice and enlargement of the penis. These effects are evidently due to hypersecretion of the male hormone, they tend to subside after excision of the tumor. Hypergenitalism with closely similar features occurs as a result of hyperfunctioning of the adrenal cortex (p. 861). Failure of development of the accessory male sex organs and of the secondary sexual characters, a condition referred to as *hypogonadism* or *eunuchoidism*, is in most cases due primarily to a pituitary disorder. Deficiency of the gonadotrophic hormone of the pituitary leads to atrophic changes in the testes, which may be greatly reduced in size, and to the suppression of their internal secretion. In other instances of eunuchoidism the testes are the primary site of disease, e.g., destructive new growths, mumps, typhoid fever, syphilis, etc. Testicular atrophy occasionally results from prolonged and severe inanition or vitamin B₁ deficiency. Hypofunctioning of the testes may commence at any time during postpuberal life as a result of pituitary disease or of any of the conditions just mentioned. In this *deferred type of eunuchoid-*

ism regressive changes, though usually of mild degree, in the accessory organs of sex and in the secondary sex characters may result. In many instances of hypogonadism in females as well as in males the excretion of gonadotrophins in the urine is increased.

THE PROSTATE GLAND

The prostate which surrounds the first inch or so of the urethra is composed of branched tubuloalveolar glands, grouped into about 20 lobules. Its stroma is fibroelastic in nature and contains many bundles of smooth muscle. The alveoli of the glands drain into some 20 ducts which discharge into the urethra. The gland is traversed by the ejaculatory ducts. Within the lumina of some of the alveoli, small concretions known as *corpora amylacea* are found especially in older men. They are composed of a calcium phosphorus compound similar in chemical structure to apatite (p. 862).

Glands lying in relation to the wall of the upper part of the urethra are present in the female. They are analogous to the male organ and are referred to as the *female prostate*.

The prostate does not appear to have any endocrine function. It secretes small amounts of fluid continuously into the urethra (from 0.5 to 2 cc. per hour in man), and much larger amounts during coitus. This occurs before the seminal vesicles discharge their secretion. Both secretions are added to the sperm, of the two the secretion of the seminal vesicles appears to be of the greater importance for the viability of the spermatozoa, or at any rate for fertilization, for this is possible after removal of the prostate but not after the loss of the seminal vesicles.

Human prostatic secretion is a thin opalescent fluid with a specific gravity of about 1.022 and containing about 7 per cent of total solids. These consist of inorganic salts, Na, K, Cl and P and about 2.5 per cent of protein, most of which is derived protein not coagulable by heat. It also contains non protein nitrogen, lipids, cholesterol, citrate, traces of glucose and the two enzymes, acid phosphatase and fibrinolysin.

The secretion of prostatic fluid is increased by stimulation of parasympathetic (sacral outflow) and sympathetic nerves (hypogastric).

Prostatic hypertrophy The cause of prostatic hypertrophy is still a matter for speculation. The theory has been advanced that the production of male hormone in excess is a factor. It is unquestionable that testosterone stimulates prostatic growth, castration causes prostatic involution. It is improbable, however, that hyperactivity of the interstitial cells is a cause of prostatic enlargement in man, secretion of male hormone tends to decrease rather than to increase with advancing years. Huggins suggests that the threshold of the prostatic tissue for the action of testosterone is

lowered as age advances to an extent which more than compensates for the diminished production of the male hormone

It has been mentioned that estrogen is excreted in the urine of the male and can be isolated from testicular tissue, it has also been shown by Zuckerman and Parkes that injections of estrogen into monkeys cause prostatic hypertrophy, fibromuscular overgrowth of the whole prostate together with epithelial stratification and distension of the uterus masculinus. Such effects can be counteracted by injections of male hormone. These facts, taken in conjunction with the observation that the concentration of the male hormone in the urine of elderly men may be reduced while the excretion of estrogen remains unchanged, have led some (de Jongh, Laqueur) to a conclusion as to the cause of prostatic hypertrophy in man based upon an imbalance between androgen and estrogen—a *diminished* production of the former and, in consequence, a relative *excess* of the latter. In support of this hypothesis de Jongh states that injections of the male hormone in cases of prostatic enlargement prevents further hypertrophy and may actually cause shrinkage of the organ. Moreover, R. A. Moore and his associates observed that prostatic tissue of the rabbit transplanted into the anterior chamber of the eye showed a greater growth response to estrogen than to male hormone administration. However, observations opposed to this theory can be cited, e.g., the absence of excessive amounts of estrogen in the blood or urine of subjects of prostatic hypertrophy, and the fact that injections of estrogen do not, apparently, aggravate the condition. Moreover, the effect of estrogen is mainly on the stroma, whereas, the male hormone stimulates the glandular element which is chiefly affected in hypertrophy of the prostate. The whole question of the causative factors in prostatic hypertrophy is rife with speculation and beset with contradictory observations and opinions. Established facts are few and difficult to obtain.

The epithelial cells of a prostatic adenocarcinoma are dependent for their growth and activity, as is the normal epithelium of the prostate, upon the male hormone. It is upon this basis that castration has been employed in the treatment of prostatic cancer, the malignant growth, in many cases, undergoing regression after operation. Since androgen activity is antagonized or neutralized by estrogens, inhibition of the cancerous growth can

also be induced by the administration of estrogenic material, e.g., stilbestrol.

The hormonal treatment of male sex disorders

The testis hormone, in accordance with the principle of hormone action in general (p. 782) does not stimulate the interstitial cells of the testes. Though highly successful results, of a temporary nature at least, have been reported following the use of the male hormone in underdevelopment or regression of the accessory male sex organs, it is too early to attempt any real appraisal of its ultimate therapeutic value in these conditions.

The gonadotrophic principle derived from pregnancy urine (p. 887) has been employed to stimulate spermatogenesis in an undeveloped testis or to restore the normal spermatogenic function which had been suppressed as a result of disease. Sterility in man is said to have been cured by this means, spermatozoa which before treatment were few and non-motile have, according to report, been increased in number and rendered actively motile after a series of injections of the anterior-pituitary-like hormone. Notable success has followed the use of the gonadotrophic hormone of pregnancy urine in cryptorchidism, descent of the testes being induced. It has also been employed in various types of hypogonadism. The gonadotrophic principle of the pituitary itself would be expected to give the best results but so far a reliable commercial preparation of this hormone has not been available.

THE THYMUS

Structure

The thymus arises from the third branchial cleft (and sometimes the fourth) on either side, each anlage going to form one of the lobes of the thymus. Each thymic lobe is composed of a number of lobules in which an outer portion or *cortex* and a central portion or *medulla* may be distinguished. The *cortex* resembles lymphoid tissue, being constituted of masses of small round cells—*thymocytes*—identical in appearance with small lymphocytes. The epithelial element of the embryonic structure is almost entirely replaced during development by the ingrowth of these lymphoid cells from the surrounding mesenchyme. A small number of elongated reticular cells are seen scattered among the lymphoid elements. The *medulla* also contains lymphocytes but in fewer numbers; the reticular cells are thus shown up more prominently and are seen to form a definite reticular stroma. Scattered throughout the medulla are round or oval elements from 30 to 100

microns in diameter known as *Hassall's corpuscles*. These bodies, which are the remnants of the original epithelial elements, are composed of cells arranged concentrically, they stain with acid dyes and therefore stand out conspicuously against the surrounding basophilic substance

Possible functions

The thymus of the infant is of relatively large size but during later childhood its weight in relation to body weight gradually decreases, little change in its absolute weight (25 to 40 grams) occurs, after the age of puberty a definite involutionary process commences. Though the involutionary changes, which consist of a reduction in the number of lymphocytes and reticular cells and their replacement by fat, are most marked during adolescence, they continue slowly throughout the rest of life. The corpuscles of Hassall disappear more slowly than the other elements.

It is admitted by most observers that the thymus serves a lymphopoietic function. Beyond this little is definitely known concerning its physiological rôle. It is enlarged in exophthalmic goiter, myasthenia gravis, adrenal insufficiency and in certain leukemic states. Thymic enlargement has also been considered to be a feature of the so called *status thymico-lymphaticus*, a condition believed to consist of hypoplasia of the vascular system, a general increase in lymphoid tissue throughout the body, and a tendency to fatal syncope. Infants and young children who have died suddenly as a result of some trifling shock or during anesthesia have been thought to be subjects of this disease, it has been thought that enlargement of the thymus was in some way responsible. An investigation carried out in 1931 by Turnbull and Young for the Medical Research Committee of Great Britain failed to substantiate this belief. In the post-mortem examination of a number of children's bodies no relationship between the size of the thymus and vascular hypoplasia to the amount of lymphoid tissue in other parts of the body was found. In subjects dying suddenly as a result of shock or during anesthesia the thymus was not larger than in subjects dying from other causes. These investigators conclude that there is "no evidence that so-called status thymico-lymphaticus has any existence as a pathological entity." The possibility has been suggested that such cases of sudden death in infants are due to failure in function of the adrenal cortex.

The great body of experimental work which has been carried out in the past in efforts to elucidate the functions of the thymus have yielded little evidence which would enable it to be classed definitely among the glands of internal secretion. There is certainly no feature of its minute structure which suggests a glandular function. Numerous investigators, nevertheless, have claimed that their findings pointed to the thymus as playing an endocrine rôle. The problem has been studied both by the use of extracts of thymus tissue and by observing the effects of extirpation of the organ. Among the functions claimed for the thymus as a result of these two lines of investigation are, the regulation of calcium metabolism and the control of skeletal growth. Defective mineralization of the bones has been described following thymectomy, others believe that this operation results in dwarfing. Gudernatsch found that feeding thymus tissue to tadpoles stimulated their growth but delayed metamorphosis. Thymectomy in pullets and pigeons is said to result in their laying eggs with uncalcified shells. Asher and his colleagues claim that thymic extracts contain a growth promoting principle which they have named "thymocrescin", this material is also said to stimulate the sex organs. The fact that thymic involution is delayed by castration but is most pronounced at the age of puberty has led several observers to the belief that the thymus is concerned in some way with sexual development.

Much of the earlier work on the thymus is very difficult to appraise, owing to the indecisive nature of the data or to the conflict between the results obtained by different workers in the field.

Anderson has made the interesting observation that rats exercised to the point of exhaustion show atrophy of the thymus and hypertrophy of the adrenal cortex. Selye has described similar changes following a variety of injurious agencies as part of what he has termed the "alarm reaction".

THE PINEAL BODY OR EPIPHYSIS

Origin and structure

This is a small gland like structure (about 10 mm long in man) somewhat resembling a pine-cone (*pinex*) in shape, situated just beneath the splenium of the corpus callosum and resting in the groove between the superior colliculi. The pineal arises as a diverticulum of the roof of the third ventricle. The cavity of the original pouch eventually disappears, the fully developed gland being composed of a solid mass of cells.

The histological features of the pineal are very diverse, the picture varying from species to species and with age. In general however, it may be said to have a pseudo-alveolar structure, the cells being arranged in masses or lobules surrounded by a highly vascular connective tissue. The cells composing these masses are of two main types: (a) *Parenchyma cells*. These are large with pale nuclei surrounded by a reticulated protoplasm containing oxyphil granules. Each cell has numerous long processes, many of which end in club-like swellings. (b) *Neuroglial cells*. These are scattered among the parenchyma cells and, as a rule, are not numerous.

Their long fibrous processes interlace with one another to form a framework for the lobule.

Involutionary changes are said to commence in the human pineal body about the seventh year. After this age laminated bodies composed of phosphates and carbonates of calcium and magnesium—the so-called “brain sand”—make their appearance.

The *functions* of the pineal body are quite unknown. Its structure has suggested an endocrine function but there is really little definite evidence from which such a conclusion may be drawn. It may be merely vestigial.

SECTION VIII THE NERVOUS SYSTEM

By N B T

(Exclusive of Chap 72)

CHAPTER 62

INTRODUCTORY THE PHYSIOLOGICAL PROPERTIES OF NERVE

THE STRUCTURE OF NERVOUS TISSUE

The structural unit of the nervous system is the *nerve cell* or *neuron*. Other elements—*neuroglial cells*—lying among the nerve cells serve as a supporting framework.

The neuroglia probably serves also as insulating material between neighboring neurons, preventing changes in electrical potential from spreading from one neuron to another. The neuroglial cells possess numerous branching processes which interlace with one another to form a dense felt work between the neurons (fig 62.1). These interstitial cells of the central nervous system are present in both the gray and white matter,

Though neurons do not multiply in the adult body and when destroyed are not replaced, certain neuroglial cells possess the power of active proliferation. This may occur to a marked degree in pathological processes.

The neuron

The neuron consists of a *body* or *soma*, and two types of process—the *dendrite* and the *axon* (axis-cylinder process, fig 62.2). The bodies of the nerve cells lie within the gray matter of the central nervous system or in outlying ganglia, e.g., posterior spinal root, cranial or sympathetic ganglia. The white matter of the brain and spinal cord and of the peripheral nerves is composed of bundles of nerve fibers. The core of each nerve fiber is formed by a process of a nerve cell. The gray matter receives a rich blood supply from the vessels of the pia mater, the blood supply to the white substance is much less profuse.

There are a great number of different types of nerve cell, those in which axon and dendrite arise by a common stem are called *unipolar*, and those in which the axon and the dendrite or dendrites spring from opposite or at least different parts of the soma are called *bipolar* or *multipolar*. The cell bodies or somata are of various sizes and forms—stellate, round, pyramidal, fusiform, etc.

By means of special staining methods four structures can be distinguished and should be especially noted in the cytoplasm or *perikaryon* of the nerve cell body, (a) *neurofibrils*, (b) *Nissl bodies* or *tigroid substance*, (c) *Golgi apparatus* and (d) *mitochondria*. The neurofibrils are very fine filaments which stream through the cytoplasm from dendrites to axon (fig 62.3), they enter the latter process and extend to its terminations. The Nissl bodies are granular masses stainable with basic dyes. They give a striped or tigroid appearance to the cell (fig 62.4A). They are absent from the region of origin of the axon and vary in size and number in accordance with the functional state of the neuron, they undergo disintegration (chromatolysis) in a fatigued or injured cell or in one whose axon has been sectioned (p 913). The internal reticular apparatus of Golgi is a coarse network seen within the cells when special methods—e.g., impregnation with silver chromate—are employed which leave the Nissl bodies and the neurofibrils invisible (fig 62.4B). It is a debatable question whether the Nissl substance is present in the living cell in the

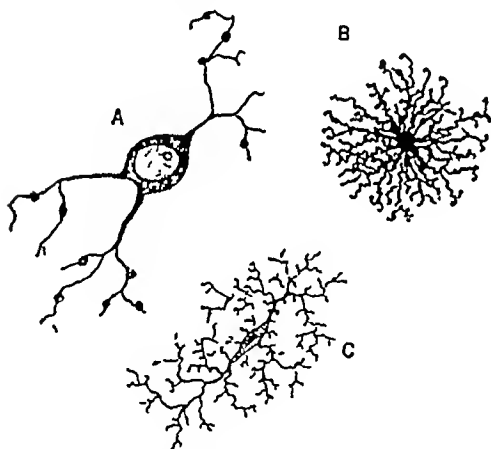


FIG 62.1. Three types of neuroglial cell. A, oligodendroglia, B, astrocyte, C, microglia.

they vary greatly in size and shape and their processes in number and arrangement. Upon a basis of these differences they have been classified into three main types—*astrocytes*, *microglia* and *oligodendroglia*. The microglia are considered by most observers to be reticulo-endothelial elements (p 105). They are migratory and phagocytic, wandering into the nervous tissue from the meninges along the blood vessels. In inflammatory processes involving the central nervous system these cells are increased in number. The oligodendroglia are believed to play a part in the formation of the myelin material which sheathes the nerve fibers.

particulate form observed, or whether the latter represents the precipitation from colloidal solution of some material—probably a nucleoprotein—by the methods of fixation employed. In either event, however, they appear to be a source of energy for the functional activity of the nerve cell.

The mitochondria are minute bodies scattered through the cytoplasm. They take the form of short rods, filaments or beads; they are not peculiar to nerve cells. The surfaces of most cell bodies are covered by a fine network—the *superficial reticulum of Golgi*. The *nucleus* of the nerve cell contains one and sometimes two nucleoli but, as a rule, no centrosome. The nucleus stains poorly, as a rule, due apparently to its paucity in chromatin. The absence of a centrosome indicates that the highly specialized nerve cell has lost its power of division. Nerve cells once destroyed are, therefore,

the neuron, the axon the discharging process, i.e., the former transmits the impulse toward, the latter away from the cell body. Nerve fibers which carry impulses to the central nervous system are termed *afferent*, those conveying impulses from the central nervous system to the periphery are called *efferent*. Purely sensory (afferent) nerves are therefore composed, strictly speaking, of dendrites, and purely motor (efferent) nerves of axons. A mixed nerve contains fibers of both types.

Structure of the nerve fiber

The axons and dendrites so long as they remain within the gray matter are simple protoplasmic extensions of the cell body, but upon entering the white

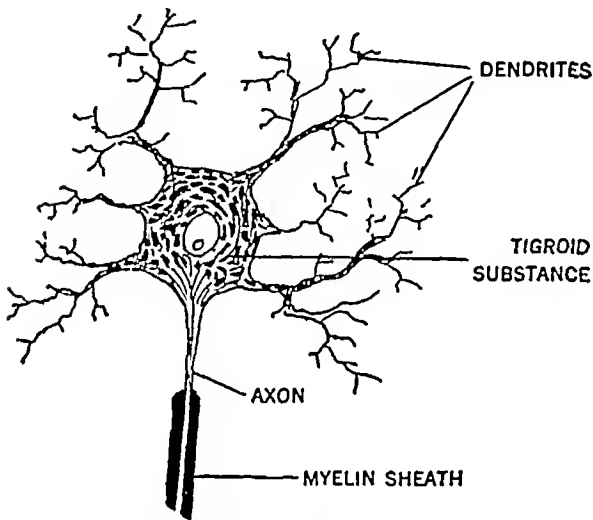


FIG 62.2 Showing different parts of the neuron

replaced merely by neuroglia, tissue quite devoid of specific nervous function.

The arrangement in series of neurons to form conducting pathways of two or more links and of variable lengths is effected by the contact (but not union) of the axon terminal of one nerve cell with the body or dendritic process of another. This functional union is called a *synapse* (ch. 64).

Though the nerve cell frequently possesses more than one dendrite the axon is single. The axon may be long and contribute to one of the tracts of the central nervous system forming the white matter, or terminate as a peripheral nerve fiber. Such cells are referred to as *Golgi I type*. In the *Golgi II type* cell, the axon is short and ends within the gray matter by making contact with another neuron. The axon arises from a small elevation on the surface of the cell body—the *axon hillock*. It may give off short collateral branches or may run as an unbranched fiber, not dividing until it has reached its destination. The dendrite is the receptive process of

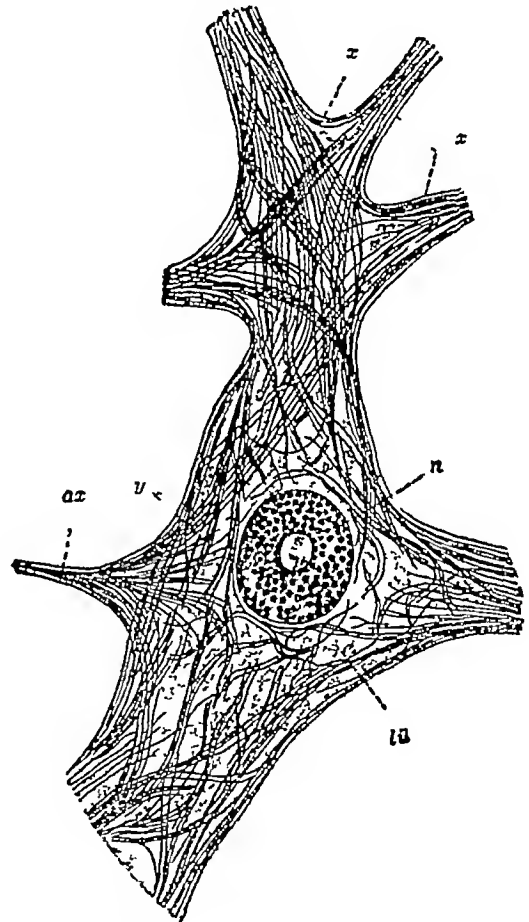


FIG 62.3 Showing neurofibrils in a cell from the anterior gray column of the human spinal cord. ax, axon; li, interfibrillar spaces; n, nucleus; x, neurofibrils passing from one dendrite to another; y, neurofibrils passing through the body of the cell (from Ranson, after Bethe and Heidenhain).

matter they become invested by a layer of lipid material called *myelin*. This covering is known as the *myelin sheath* (fig 62.2). In the peripheral nerves, but not in the central nervous system, the myelin sheath is enclosed in turn by a nucleated membrane, the *neurilemma* or *sheath of Schwann*. In peripheral nerve fibers

the myelin sheath is interrupted at regular intervals. The neurilemma dips into the gaps so formed to give the appearance of evenly spaced constrictions known as the *nodes of Ranvier*. Each section of the neurilemma between the nodes of Ranvier consists of a single cell—the *cell of Schwann*. Toward its termination the nerve fiber loses its myelin covering, being then clothed simply by the neurilemma. The latter is finally lost, the fiber terminating as a naked axis cylinder which may or may not become enclosed within some form of specialized end organ (ch 63). The processes arising from sympathetic nerve cells (post ganglionic fibers) are devoid of a myelin covering. They are invested simply by a sheath of Schwann and are therefore called *amyelinated* or *nonmedullated* fibers. There is a direct relationship between the diameter of nerve fibers and the speed at which they conduct nerve impulses (p 925).

Myelination of fiber tracts in the central nervous system

The nerve fibers in the various conducting pathways receive their myelin sheaths at different ages and it is generally believed that the myelination of a given tract and the time at which it commences to function coincide. The sensory tracts become myelinated first, those of the posterior columns of the spinal cord between the fourth and fifth months of fetal life (human). The spinocerebellar tracts are myelinated later and the motor paths, e.g., corticospinal (pyramidal) tracts do not commence to receive their myelin sheaths until the second month of life and are not completely myelinated until about the second year, or about the time when the child has learned to walk. The fibers of association paths, for the most part, myelinate at still later dates.

Neurobiotaxis

Kappers explains the development of conducting pathways in the central nervous system with the theory that the body and dendrites of a nerve cell move, as a result of some attractive force, toward the point from where its stimuli come. Its axon lengthens in the opposite direction. Kappers terms this process *neurobiotaxis*. The nature of the attractive force is unknown, though it has been suggested that neurobiotaxis is a galvanotropic effect, the point in the nervous tissue in receipt of stimuli being electrically negative, presumably, to its surroundings. The theory of neurobiotaxis offers a ready explanation for the development, as a result of experience and training, of the innumerable reflex arcs within the central nervous system and of the multiplication of association pathways. The close relation of the eye muscle nuclei to the medial longitudinal bundle, through which impulses from numerous sources are received and the peculiar course of the fibers of the facial nerve within the pons, are other examples which are believed to illustrate the

operation of this principle. The course taken by the facial fibers, which arch over the abducens nucleus, is attributed to the migration of the facial nucleus toward the tractus solitarius and the sensory nucleus of the

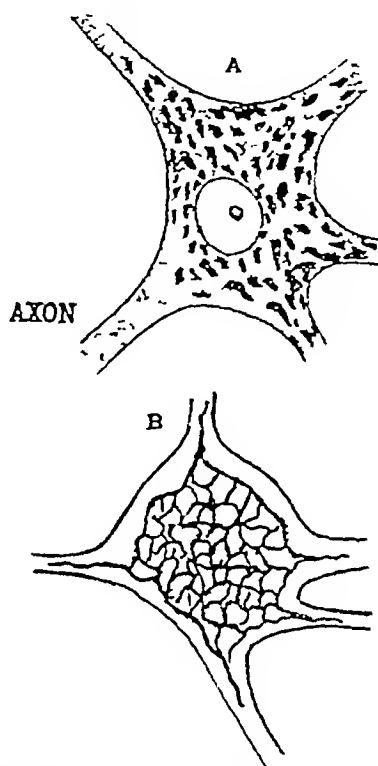


FIG 62.4 A, showing Nissl bodies, B, internal apparatus of Golgi

trigeminal, from both of which the facial neurons receive impulses

NERVE DEGENERATION AND REGENERATION

When a nerve fiber is divided, the portion peripheral to the point of section, being separated from the body of the cell, undergoes degenerative changes. It is generally believed that the degenerative process does not start first at the point of section and progress to the periphery but that all parts of the nerve fiber even to its finest terminals are involved virtually simultaneously. The first change is noted in the neurofibrils which become tortuous and show irregular thickenings. The myelin sheath next becomes swollen and breaks up into small ovoid segments (fig 62.5). Later, decomposition of the myelin occurs: droplets of cephalin and lecithin appear, derivatives of the latter—choline, glycerophosphoric acid and unsaturated fatty acids—can be detected. There is

a progressive and rather rapid fall in the acetylcholine (ch 72), cholinesterase and thiamine contents of the degenerating nerve. The fatty acids are responsible for the brown staining of the degenerated nerve when treated with osmic acid. The nuclei of the neurilemma proliferate, and the cytoplasm becomes swollen and vacuolated. The neurofibrils undergo fragmentation, they and the protoplasm surrounding them finally undergo complete disintegration, the debris intermingling with the material of the disorganized myelin sheath.

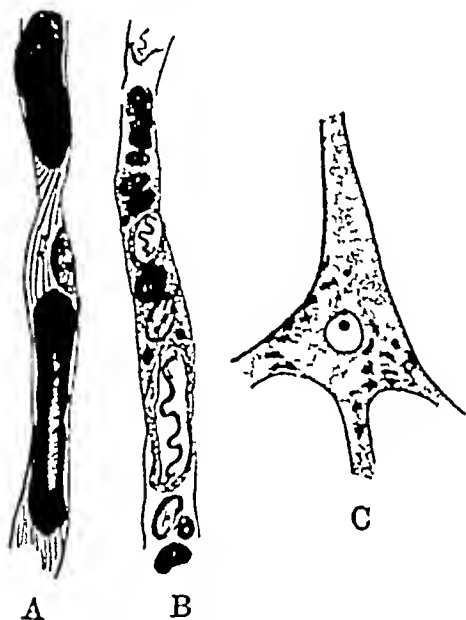


FIG 62.5 Degenerating nerve stained with osmic acid. A, shows appearance of distal segment of nerve fiber 2 days after section, note large masses of myelin derived from medullary sheath, B, 5 days after section, smaller myelin particles together with droplets of fatty acids and fragmented neurofibrils, C, retrograde degeneration in cell body, disintegration of Nissl bodies.

Subsequently the degenerated materials disappear, and all that remains of the nerve fiber is the empty neurilemma tube. The foregoing changes were first observed by Waller and are spoken of as *Wallerian degeneration*. Alterations which were not recognized by Waller also occur in the neuron on the proximal side of the section (*retrograde degeneration*). The nerve fiber as far centrally as the first node of Ranvier shows changes similar in nature to those just described. In the cell body itself, swelling of the cytoplasm and nucleus occurs and the Nissl granules undergo disintegration (*chromatolysis*). Atrophy of the cell body may ultimately result.

Regeneration. This occurs in peripheral nerve fibers but not in those of the central nervous

system, the presence of a neurilemma being essential for the process. Section of the sensory root of the trigeminal nerve, for example, for the relief of severe neuralgia of the face, is not followed by regeneration of the fibers within the brain. Nor does regeneration of the optic nerve, whose fibers do not possess neurilemma sheaths, occur.

Regeneration is accomplished by a downgrowth of the neurofibrils from the proximal segment of the cut nerve. These enter the empty neurilemma sheath of the distal segment and in time traverse its entire length. The rate of growth in man has been variously estimated (from 1.0 to 4.5 mm per day), it is influenced by many factors, namely, the nerve involved, whether there is complete severance of the nerve (*neurotmesis*), the distance separating the sections, and whether the axons alone are crushed, the supporting structure remaining intact (*axonotmesis*). The cells of the neurilemma form a protoplasmic matrix for the sprouting neurofibrils. The neurilemma of the lower segment grows upwards, and if the gap is not too long joins that of the upper segment which has formed around the developing axis cylinder. The myelin sheath is developed some time later. The young axis cylinders may traverse the tissues for relatively long distances in order to enter the neurilemma of the lower segment. The force responsible for this remarkable phenomenon (*neurotropism*) is unknown. The attraction may be due to the liberation of chemical substances by the proliferating neurilemma cells of the lower segment. If the gap separating the two segments is wide and occupied by scar tissue, an effectual block may be offered to the regenerating fibers, these nevertheless continue to grow in a tangled, curled fashion and may produce a localized mass of tissue composed of nerve fibers embedded in connective tissue (*neuroma*). In those cases of nerve injury in which, as a result of pressure or crushing, the axis cylinder is interrupted without division of the neurilemma, the conditions for regeneration are the most favorable.

The remarkable rapidity of regeneration of preganglionic sympathetic fibers has been demonstrated by Hamovici and Hodes. In a three-stage operation upon cats large sections of the sympathetic chains below their cervical portions were excised. Stimulation of the cephalic end of a cervical sympathetic stump as early as 54 days after the last stage of the operation caused pupillary dilatation and retraction of the nictitating membrane.

The regenerating fibers of one nerve will grow into the sheath of the lower segment of another, though less readily than into its own sheath, and even the fibrils of a sensory nerve will grow into the distal segment of a motor nerve or vice versa. The central stump of one nerve will not, however, grow into the central stump of another. The proximal end of the hypoglossal or spinal accessory, for example, has on many occasions been anastomosed to the distal end of a paralyzed facial nerve with a successful functional result. Cannon, Binger and Fitz united the phrenic nerve in cats to the peripheral (upper) end of the cut cervical sympathetic low in the neck and later observed a rise in metabolism, nervous excitability and increased heart action which were attributed to stimulation of the thyroid by impulses discharged from the respiratory center. This experiment was repeated by Horrax, and more recently by Friedgood and Cannon, with essentially the same result. The effect is probably not a direct one upon the thyroid but is due to stimulation of the pituitary and the release of its thyrotrophic hormone. Balance and Duel, also experimenting with cats, anastomosed the central end of the hypoglossal to the peripheral cut end of the cervical sympathetic, and reported that the normal pupillary reactions were restored. Restoration of function does not follow the union of a motor with a sensory nerve, nor of a cholinergic with an adrenergic nerve (ch 72).

It has been shown by Tello in animals and by Duel in man that when a nerve has been severed a much better functional result is obtained if from 2 to 3 weeks are allowed to elapse before the divided nerve is sutured. The reason for this is that the degenerated material in the distal segment has had time to be cleared away, thus, an unobstructed neurilemma tube is left for the downward growth of the neurofibrils of the upper segment. Otherwise, apparently, some of the neurofibrils upon meeting the degenerated debris are diverted from their course.

Degeneration of the facial nerve in the facial canal as a result of middle ear disease (Bell's palsy, ch 66), trauma, etc., has been treated with outstanding success by Duel. The affected portion of the nerve is excised and a graft, constituted of a section of the anterior femoral cutaneous nerve, is used to fill the gap. For the reason just given the latter nerve is sectioned and left in place for 2 or 3 weeks before it is employed in the grafting operation.

Reaction of degeneration

A motor nerve may be stimulated through the skin by either the faradic (interrupted) or the galvanic (constant) current. The muscle contracts so long as the faradic current flows but only during the make (closure) or break (opening) of the galvanic current, for a given strength of current closure shocks are more effective than opening shocks. In testing the reactions one electrode is placed upon the skin of some indifferent part of the body (indifferent electrode) and another smaller electrode (stimulating electrode) is placed upon the skin overlying the nerve trunk or muscle which it is desired to stimulate. In the latter instance the nerve terminals within the muscle are stimulated, not the muscle fibers themselves. The stimulating electrode is applied over that part of the muscle which gives a response with the least strength of current, this is spoken of as the motor point and corresponds to the point of entrance into the muscle of the motor nerve. When the stimulating electrode is attached to the positive pole of the battery it is called the anode. The current enters the muscle by this electrode and leaves the body by the indifferent electrode which is called the cathode. When the galvanic current is reversed, the stimulating electrode becomes the cathode and the indifferent electrode the anode. Normally the least strength of current is required when the cathode overlies the muscle and the circuit is closed. The response which follows such a shock is called the cathodal closing contraction (abbreviated CCC). The next most easily elicited response is the anodal closing contraction (ACC), i.e., when the stimulating electrode in the anode and the current is closed. The anodal opening contraction (AOC) is less easily elicited than the last, and the cathodal opening contraction (COC) requires the strongest current of all. The four reactions in the order of the strength of current required for their elicitation may be expressed thus

$$CCC < ACC < AOC < COC$$

The normal relationship between current strength and these responses is also shown in the following table

STRENGTH OF CURRENT	REACTION
Weak	CCC
Medium	CCC and ACC
Strong	CCC, ACC, AOC.
Very strong	CC tetanus, ACC, AOC and COC

The investigation of these reactions is of considerable value in the detection of degeneration of a motor nerve and in the diagnosis of a lower from an upper neuron lesion (ch 66). Lengthening of the chronaxie of a nerve occurs during the earlier stages of its degeneration, it therefore fails to respond to the brief shocks of

faradic stimulation but a sluggish contraction of the muscle follows stimulation with the galvanic current. Later the nerve becomes quite incapable of being excited by either the faradic or the galvanic current. When the nerve terminals have degenerated the chronaxie of the muscle is lengthened and the faradic current applied directly to the muscle causes no response, the muscle responds, however, though in an abnormal manner, to the galvanic current, the fibers being directly stimulated. The response of the muscle shows the following features (a) Sluggishness of contraction and relaxation (b) The response is elicited with a weaker current than normally, i.e., the muscle is hyperexcitable to the galvanic current. (c) The anodal closing contraction (A C C) may often be elicited more readily, i.e., with a weaker current, than the cathodal closing contraction (C C C). Thus, $A C C < C C C$

These changes in the electrical responses of nerve and muscle, namely, loss of excitability of nerve to faradic and galvanic stimulation, failure of the muscle to respond to faradic stimulation and the abnormalities of the reaction to galvanic stimulation just described constitute the complete reaction of degeneration (C R D). When the nerve is still capable of being excited by galvanic but not by faradic stimulation, but the muscle responds weakly or not at all to the faradic current, the reaction of degeneration is said to be incomplete—partial reaction of degeneration (P R D).

THE PHYSIOLOGICAL PROPERTIES OF THE NERVE FIBER

EXCITABILITY AND CONDUCTIVITY

The nerve fiber may be stimulated electrically, thermally, chemically or mechanically. Any one of these types of stimulus causes a change in the nerve at the point of stimulation which may be termed the local excitatory state. If this attains a certain value a wave of excitation is transmitted along the nerve fiber. The propagated disturbance is referred to as the nerve impulse, and its passage from point to point along the fiber as conduction.

THE CHARACTERS OF A STIMULUS

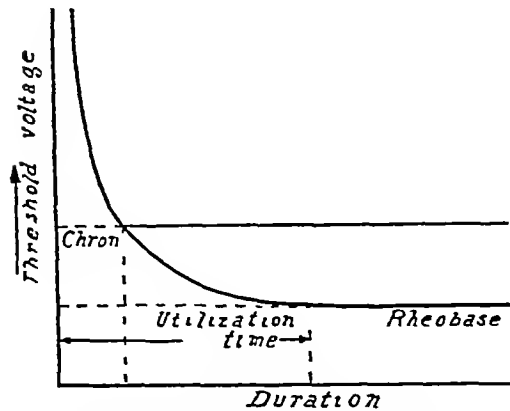
A stimulus may be defined as any change in the environment of a tissue which causes it to react. In experimental work the electric current is usually employed as the stimulus on account of its convenience and the accuracy with which it can be measured. It also leaves the tissue undamaged. In order to induce a local excitatory state in a nerve of sufficient value to set up an impulse, a stimulus must answer the following requirements as to (a) strength, (b) duration, and (c) rate of change in intensity

(a) Intensity or strength When electrical stimulation is employed the intensity corresponds, practically, to the voltage. It is well known that while a strong stimulus will cause a response, a weaker one may fail to do so. The stimulus which possesses just sufficient strength, and no more, to set up an impulse is said to have an intensity of threshold (liminal) value. Stimuli of less strength than this are called subthreshold (subliminal) or subminimal. A subthreshold stimulus does not, however, leave the nerve unaffected for, if a second stimulus also of subthreshold strength, or a series of such stimuli, be sent into the nerve an impulse is set up. The first stimulus is believed to cause a local excitatory state, or "local potential" (Hill), which, though of too low a value to set up an impulse, can be built up to the requisite value by a second, or a series of subthreshold stimuli, applied at short intervals. The phenomenon is spoken of as the summation of inadequate stimuli.

(b) Duration When a constant current is passed through nerve or muscle, excitation occurs only at the moment that the circuit is made or broken. For this reason the duration of the current flow was believed in the past to bear no relationship to its ability to excite. The strength of the current and the rate at which this developed to a maximum ((a) and (c) above) were considered to be the sole factors determining the effectiveness of the stimulus. Though it is true that no physiological effect is produced in the nerve by prolonging the period of current flow indefinitely, nevertheless the current must flow for a certain period in order to be effective. For example, a high frequency alternating current of high voltage (tesla) does not excite (though it causes heating of the tissues). For the same reason a faradic current of very high frequency may fail to excite, whereas one of lower frequency is effective. Presumably then, a certain period of time is required for the current to bring about those changes in the nerve upon which the excitatory state depends. This period throughout which there is a reciprocal relationship between strength of current and the duration of its flow is called the serviceable or utilization time (temps utile). It is measured in thousandths of a second. When, for instance, a current of a certain voltage and having a duration of say 3 sigmas (σ) or milliseconds (m sec)¹ just fails to excite the tissue, then either increasing the voltage or prolonging the duration of the current causes a response. Or, if a

¹ A sigma or millisecond (m sec) = $\frac{1}{1000}$ second

current of a certain voltage and duration is just capable of exciting the tissue, reducing either of these factors renders the stimulus ineffective. The length of the utilization time varies in different tissues, but for the frog's gastrocnemius it is about 30 sigmas. Beyond the utilization period no relationship exists between the intensity and the dura-



CHRONAXIE It is evident from the foregoing paragraphs that the excitability of a tissue may be determined in one or other of two ways (a) by ascertaining the minimal strength of a current which when allowed to flow for an unlimited period will excite (intensity threshold), or (b) by measuring the minimal time during which a current of a standard strength must flow in order to excite (duration threshold) Lapicque employs the latter or, as he has termed it, the *time characteristic* or *chronaxie*, considering it a more accurate measure of excitability. A current having a strength twice the rheobase is employed, and the minimal duration of flow necessary for excitation is measured. The *chronaxie* may therefore be briefly defined as the shortest duration of a current necessary for excitation when its strength is twice the rheobase.

Determinations of chronaxie may be carried out by the use of (a) a constant current which by means of a special type of rheotome is allowed to flow through the tissue for a very brief accurately measured period, (b) condenser discharges. The duration of a condenser discharge is proportional to the capacity (C) of the

FIG 62.6 Strength-duration curve, chron., chronaxie.

TABLE 86

SPECIES	TISSUE	CHRONAXIE	TESTS UTILE	DURATION OF CONTRACTION
		m.sec	m.sec.	seconds
Dog	Ventricular muscle	2 0	—	0 16
	Ventricular muscle during vagal stimulation	0 8	—	—
	Bundle of His	6 0	—	—
	Gastrocnemius	0 3	3 0	0 1
Frog	Stomach	100	—	15-20
	Sciatic nerve (conduction 30 m per second)	0 3	3 0	—
Turtle	Ventricular muscle	9 0	80	—
	Leg muscles	1-2	—	1 0
	Flexors of thigh	0 10-0 16	—	—
	Extensors of thigh	0 44-0 72	—	—
Man	Flexors of arm	0 08-0 16	—	—
	Extensors of arm	0 16-0 32	—	—
	Retina	1 2 -1 8	—	—
	Vestibular N	14 -22	—	—
	Taste buds (tip of tongue)	1 6	—	—

tion of the current. Therefore when the current is of indefinite duration its effectiveness depends entirely upon its intensity. The intensity (voltage) of a current which when allowed to flow for an indefinitely long period is just capable of exciting the tissue (intensity threshold) is called the *rheobase*. The relationship between strength and duration of current is shown in figure 62.6

condenser, provided the resistance (R) in the discharge circuit remains constant. By charging condensers of different capacities (measured in farads) to different potentials, currents of any desired duration or voltage can be employed. In order to render any variation in the resistance of different tissues negligible a high resistance (15,000 to 20,000 ohms) is introduced into the discharge circuit. In determining the chronaxie by means of condenser discharges the current of

minimal voltage required for excitation is first determined, a very high capacity, i.e., one with a long discharge, being used. The threshold voltage (rheobase) is then doubled and the lowest capacity is found (shortest duration of discharge) which will produce a response at the higher voltage. The result is obtained in microfarads. Since the time, or chronaxie, is proportional to the product of the capacity and the resistance in the discharge circuit, the result can be converted into seconds by the following formula:

$$\text{Chronaxie} = C \text{ (farads)} \times R \text{ (ohms)} \times K \\ (K = 0.37)$$

Rapidly reacting tissues have a shorter chronaxie than the more slowly acting. The chronaxies of smooth muscle and its nerves are longer than those of skeletal muscle and somatic nerves. The tissues of cold-blooded animals have, in general, longer chronaxies than the tissues of higher forms. Flexor muscles have shorter chronaxies than extensors and the chronaxies of the more rapidly acting white muscles are shorter than those of the red. Nerve fibers of larger diameter and of more rapid conduction respond to shorter durations of current than do the thinner and slower fibers. The chronaxies of sensory nerves are in general about the same as those of the corresponding motor nerves. The chronaxie of the ventricular muscle, contrary to expectation, is shortened by vagus stimulation, or by drugs such as acetylcholine which slow the cardiac rate, and lengthened by accelerator stimulation or by drugs, e.g., atropine and adrenaline, which cause acceleration. The chronaxie of the junctional tissues is about three times longer than that of auricular or ventricular muscle. The chronaxie of the gastric muscle is also shortened by vagal stimulation. The chronaxie is lengthened by cold and shortened by a rise in temperature. Stretching cardiac or smooth muscle reduces the chronaxie, and that of skeletal muscle is lengthened by fatigue and shortened by adrenaline. The utilization time (temps utile) of a tissue varies with its chronaxie, being about 10 times the value of the latter. Thus the chronaxie of the frog's gastrocnemius is 0.3 m sec, the effective period is 3.0 m sec.

In table 86 are compared the values for chronaxies and utilization periods for different tissues.

The theory of isochronism. Lapique maintains that normally the chronaxie of a muscle and that of its nerve are of the same order—the value of one not being greater than that of the other by more than 100 per cent. He claims that this so called isochronism is essential for the transmission of the impulse from nerve

to muscle. A chronaxie difference greater than 100 per cent between a muscle and its nerve or between any other two successive parts of the conducting pathway (i.e., between neurons), is termed heterochronism. It acts as a block to the passage of the impulse. Certain muscular poisons act, it is supposed, by inducing heterochronism. Curare, for example, which does not paralyse either the muscle or its nerve was previously supposed to act upon a special "receptive substance" situated between the nerve ending and the muscle fiber. According to Lapique, the drug acts by lengthening the chronaxie of the muscle but leaving that of the nerve unaffected, i.e., heterochronism is induced. In support of the theory, it is pointed out that when a nerve-muscle preparation is treated with either veratrin or strychnine the muscle can no longer be stimulated through its nerve. Veratrin shortens the chronaxie of the muscle, strychnine that of the nerve. In either case heterochronism is induced. When, on the other hand, the preparation is treated with the two drugs simultaneously the impulse is not blocked since, the chronaxies of both nerve and muscle being shortened, isochronism is maintained. The onset of muscular fatigue is attributed to the development of heterochronism between the muscle and its motor nerve.

The theory of isochronism has been criticised by Rushton, to whose article the reader is referred, this observer was unable to confirm Lapique's finding regarding the actions of curare, veratrine and strychnine. At the best the theory appears to be only an approximation and strict isochronism is not a necessary condition for neuromuscular transmission.

Chronaxies in the human subject have been investigated under various physiological conditions and in many pathological states by Bourguignon and others. Condenser discharges are employed, an indifferent electrode being placed upon the skin over the sternum and the stimulating electrode, as the cathode (p. 914), upon the skin overlying the nerve or muscle to be investigated. The chronaxies in the new-born child are about 10 times longer than those of the adult. The longer chronaxies are in agreement with the much slower movements of the infant. Heterochronism exists between the muscles and their nerves until about the 15th month (walking age). The lengthening of the chronaxie of degenerating nerve and its muscle has already been mentioned (p. 914). In diseases of the central nervous system all muscles which are the seat of paralysis show lengthened chronaxies and in progressive nervous disease those groups of muscles of similar chronaxies are usually attacked more or less simultaneously.

(c) *Electrotonus, electrotonic currents, the rate of change in the intensity of the stimulus*
—accommodation in nerve

ELECTROTONUS. It was shown many years ago (1859) by Pflüger that the passage of a constant electric

current through a nerve altered the properties of the nerve in respect to *excitability*, *conductivity*, and *electrical state*. The effects are produced both at the *anode* (positive pole) and at the *cathode*, and are referred to as *electrotonus*, the anodal effect is called specifically *anelectrotonus*, while the effect at the cathode is called *catelectrotonus*.

The effects occur in sensory as well as in motor nerves, but they can be demonstrated most conveniently in a motor nerve attached to its muscle (fig 62.7). If the electrodes are placed upon the nerve with the cathode nearer the muscle, i.e., the current is *descending*, and the nerve then stimulated near the cathode with induction shocks which, ordinarily, are sub-

the indifferent region is nearer the cathode than the anode, i.e., anelectrotonus extends farther into the interpolar region than does catelectrotonus, with weak currents it is nearer the anode (fig 62.8).

The effect of the constant current upon *conduction* can be shown by stimulating the nerve in the interpolar region. When the anode is nearer the muscle, anelectrotonic depression of conductivity tends to block the impulse from reaching the muscle, complete block is produced by a strong current. There is no interference with the passage of the impulse when the electrodes are reversed.

The stimulating effects of making and breaking a constant current vary with the strength of the current. With a current of moderate strength stimulation of the nerve occurs at the cathode when the current is "made" (or closed) and at the anode when it is "broken" (or opened), for excitation is increased at the cathode in the one instance and at the anode in the other. The cathodal

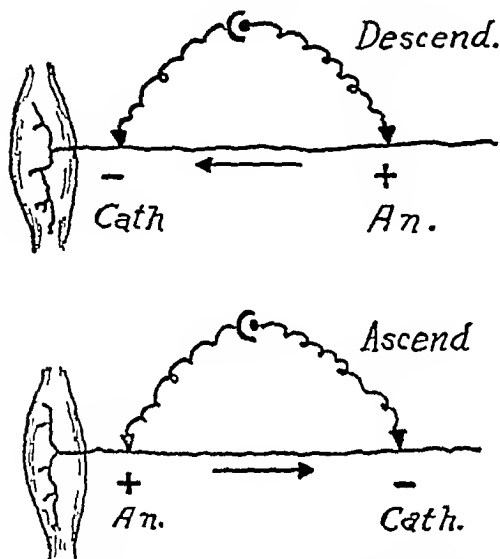


FIG 62.7 Diagram illustrating the arrangement for demonstrating anelectrotonus and catelectrotonus. With a strong descending current (upper figure), break excitation at the anode is blocked at the cathode, and with a strong ascending current make excitation at the cathode is blocked at the anode. An., anode, cath., cathode.

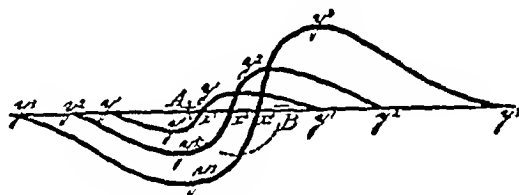


FIG 62.8 Illustrating altered excitability of nerve during the passage through it of a weak (y^1), medium (y^2) and strong (y^3) constant (polarizing) current. A, anode, B, cathode. The horizontal line represents the normal excitability of the nerve, the curves above and below this line indicate increased and decreased excitability, respectively, of the nerve. The crossings of the horizontal line by the curves, x^1 , x^2 and x^3 , indicate the points where anelectrotonic and catelectrotonic effects neutralize one another, i.e., they represent in different points in the interpolar region A-B (From Lombard after Pfleger).

maximal, the muscle gives a maximal contraction owing to the increased excitability of the nerve. The excitability at the anode can be shown in a similar way to be reduced. When the electrodes are reversed, the anode being the nearer the muscle and the current *ascending*, the nerve again shows increased excitability in the region of the cathode and reduced excitability in the anodal region. When the current is interrupted, opposite effects are produced before the nerve returns to its normal state, namely, a rise in excitability at the anode and a depression of excitability at the cathode.

The anelectrotonic and catelectrotonic effects spread along the nerve on either side of the electrodes. At a certain indifferent point between the poles one effect, as it were, neutralizes the other. With strong currents

"make" effect is always greater than the anodal "break" effect, so that, with weak currents the excitation at the anode does not occur, a response is obtained only upon making the current. With very strong descending currents, a response is obtained only on the "make" because on the "break" the impulses set up at the anode are blocked by catelectrotonic depression near the muscle. With strong ascending currents excitation occurs only on the "break" because on the "make" the impulses are blocked by anelectrotonus depression near the muscle (see table 87 and fig 62.8).

Electrotonic currents. During the passage of a constant current through a nerve the resistance of its membranes causes the current to spread laterally along their surfaces as well as toward the core of the nerve. The change in the electrical state within the nerve is thus diffused for a distance on either side of the electrode (fig 62.9). If a galvanometer be led off from the outer side of either electrode, i.e., in the extrapolar region, a

current flows through the instrument in the same direction as the constant current. The current flows toward the anode and away from the cathode (and from anode to cathode in the interpolar region). These *electrotonic currents* are due to changes in polarization of the membranes or interfaces within the nerve structure by the constant (polarizing) current. When the polarizing current is broken the electrotonic currents flow in the opposite direction. The production of electrotonic currents by a constant current can also be shown in an artificial moist conductor model consisting of a conducting wire running down the center of a glass tube filled with saline. The changes in the electrical state of the nerve are the cause of catelectrotonus and anelectrotonus described in the preceding section.

When the current is “made”, and the “local potential” at the cathode, which may be designated V ,

for the return of V to V_0 , it has a value of around 0.35 msec for frog’s nerve and is related to the chronaxie which equals $0.693 \times K$. In order for excitation to result a certain rate of change in the intensity of the stimulus is required. The nerve “accommodates” itself to the stimulus if the change in intensity of the latter is not sufficiently rapid. A stronger stimulus is therefore required for excitation when the rate of rise in intensity is slow than when it is rapid. Hill has introduced a second time factor which he calls the *time constant of accommodation* (λ). If the critical value of V for excitation, i.e., the threshold, designated U , and substituting U and U_0 for V and V_0 , and λ for K , in the above formula, we have $dV/dt = (U - U_0)/\lambda$. For the calculation of the time constant of accommodation (λ) the reader is referred to the original paper.

The time constant of accommodation is altered by several conditions, the most notable being changes in calcium ion concentration. An increase in concentration of ionic calcium reduces the value of λ , that is, the neuromuscular tissues accommodate more readily, which means that more rapid changes in stimulus strength are required for excitation. Reduction in ionized Ca, as in tetanic states, exerts the reverse effect—increased value of λ , and reduced accommodation of the neuromuscular tissues which are in consequence excited by relatively slow rates of change in the strength of the stimulating current (see p. 908). If the calcium ion concentration is sufficiently low there may be a complete absence of the phenomenon of “accommodation.”

TABLE 87

	ASCENDING CURRENT		DESCENDING CURRENT	
	Mak-ing	Break-ing	Mak-ing	Break-ing
Very weak currents	C	O	C	O
Moderate “	C	C	C	C
Very strong “	O	C	C	O

C = contraction, O = failure to contract.

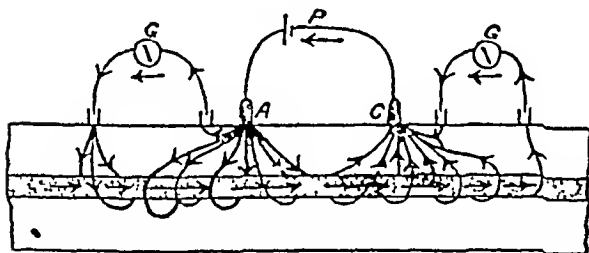


FIG 62.9 Electrotonic currents. P, constant current, A, anode, C, cathode, G, galvanometer (After Schäfer)

reaches a certain value, excitation of the nerve occurs. The critical value of V at which the nerve is excited, i.e., the threshold of excitation, may be designated U . The decay in the “local or cathode potential” upon cessation of the current, i.e., the rate of return of V to its original or resting value V_0 ,² is expressed by the formula $dV/dt = (V - V_0)/K$. K is the time constant

² The reason for measuring the decline of V to V_0 is given by Hill in the following words: “We might have defined the time constant in terms of the rate of change of V when a current is put into the nerve. This would have required the choice of a particular form of current. There would have been no difficulty, but it might have given the impression that K has something to do with the current, it has not, it is a property of the nerve itself, it expresses the time factor in the tendency of the nerve to revert to its resting condition.”

THE NERVE IMPULSE

The nerve impulse is a self-propagated disturbance. That is to say, the energy for the transmission of the impulse is derived from the nerve fiber over which it passes. Nervous conduction therefore depends upon the state of the fiber at successive points reached by the impulse. The impulse resembles a spark traveling actively along a train of gunpowder rather than a wave transmitted passively through air or water. In either of the latter two instances in contrast to the first, the energy is derived from a source other than the medium through which the wave travels and the force and amplitude of the wave become gradually reduced with distance. To carry further the analogy drawn between the impulse and the burning train of gunpowder—if a section of the powder fuse is dampened in advance of the spark, the latter becomes less intense as it passes through the dampened section, and travels more slowly. Upon reaching a succeeding dry portion the spark flares up again to its previous intensity and velocity, and, so long as the powder remains dry, is transmitted without change to the end of the

fuse In a comparable way, if the activity of a segment of nerve is depressed by treatment in a chamber with a narcotic (alcohol or ether vapor) the impulse undergoes a reduction in amplitude and velocity in its passage through the narcotized region, but upon reaching the untreated nerve beyond, regains its original value, and is transmitted unchanged to the termination of the nerve. The question then arises whether the strength of the impulse in its passage through the narcotized section of nerve is suddenly reduced or whether the reduction is progressive. In other words, would the impulse suffer a greater reduction if it were made to traverse a long section of narcotized nerve than if it passed through a short stretch? Until recently it was thought that a gradual or progressive impairment of conduction, i.e., *conduction with a decrement*, occurred. The mode of propagation of the impulse over narcotized nerve was therefore supposed to be essentially different from that along a normal nerve. If, for example, the narcotized section were long enough, complete extinction of the impulse would result, in this region, therefore, the impulse would resemble a wave transmitted through air or water. From his experiments, in which the long nerve of the Japanese toad was employed, Kato could obtain no evidence that there was a progressive decline in the strength of the impulse. He came to the conclusion that the impulse suffered instantaneous reduction upon entering the region of narcosis and underwent no further reduction during its transmission along the narcotized section. Conduction was therefore *decrementless*. Kato's findings have been confirmed by Davis, Forbes, Brunswick and Hopkins (see fig. 62 10) but it is pointed out that on theoretical grounds there must exist a short transitional portion of the nerve at the junction of the normal and the narcotized section, 7 mm or less in length, where progressive decline in the impulse occurs. Beyond this the impulse is conducted throughout the length of the narcotized section without further reduction.

Conduction rates

The velocity of the nerve impulse varies in different nerve fibers in accordance with their diameters, the thicker fibers conducting more rapidly than those of smaller diameter. In the large motor fibers of the mammal the rate is from 80 to 100 meters per second. Sensory nerves of the skin being of smaller diameter have slower

conduction rates. Non medullated fibers conduct more slowly than medullated. Some of the fibers subserving pain sensation and those of the sympathetic nervous system have a very slow conduction rate (see also p. 925).

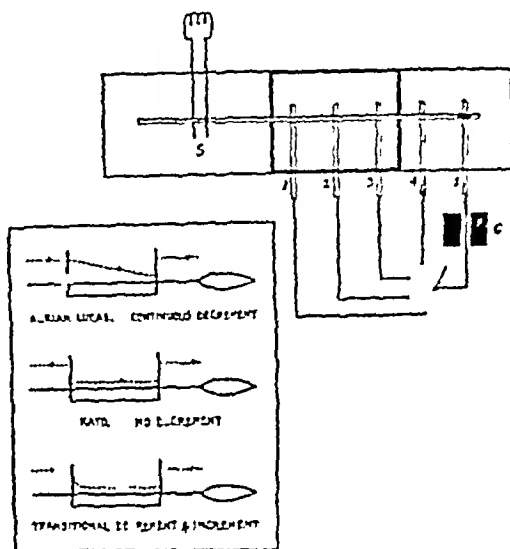


FIG. 62 10 Upper drawing, diagram of nerve (peroneal of cat) in narcotizing chamber to illustrate "set up" of an experiment on conduction of impulses through narcotized nerve. S, stimulating electrodes; Lerdoff electrodes at 1, 2, and 3 within the chamber and 4 outside. Circuit completed through an indifferent electrode at 5 G, galvanometer for recording action currents. Nerve in chamber exposed to alcohol vapor. The results of this experiment gave no evidence of conduction with a decrement. The action currents from all three leads within the chamber were reduced to practically the same degree. Lower drawing contrasts the different views advanced regarding conduction of the impulse through a narcotized region of nerve. It will be noted that in all three the impulse regains its full value upon entering healthy nerve (from Davis, Forbes, Brunswick and Hopkins).

The following table from Hill gives the conduction rates in the nerves of several different animals.

Medullated nerve, mammal, 37°C, about 120 m/sec
Medullated nerve, dogfish, 20°C, about 35 m/sec
Medullated nerve, frog, 20°C, about 30 m/sec.
Non medullated nerve, crab, 22°C, and 15 m/sec
Non medullated nerve, mammal, 37°C, about 1 m/sec
Non medullated nerve, olfactory of pike, 20°C, 0.2 m/sec
Non medullated nerve, in fishing filament of <i>Physalia</i> , 26°C, average 0.12 m/sec.
Non-medullated nerve, in Anadon, 0.05 m/sec
Compare the velocity of sound in air at 0°C, 331 m/sec.

By an indirect method of measurement Carmichael and his associates found the rates of conduction in various human postganglionic sympathetic nerves to be from 0.85 to 2.30 meters per second. The lower figures were obtained for the nerves of the leg, the higher ones for the nerves of the chest.

The "all or none" principle

A stimulus which is just capable of exciting a nerve fiber (threshold stimulus) sets up an impulse which is no different from one set up by a much stronger stimulus. The impulse set up by the weak stimulus is conducted just as rapidly and is just as strong, when judged by the action current developed or the mechanical response of the muscle, as one set up by the strong stimulus. Briefly, the propagated disturbance set up in a single nerve fiber cannot be graded by grading the intensity or duration of the stimulus—the nerve fiber gives a maximal response or none at all. To make use again of the train of gunpowder analogy—the flame of a match applied to the powder fuse will start a traveling spark no less intense than one started by the flame of a torch. The restoration of the strength of the impulse to its original value after passing from a narcotized region into healthy nerve (p. 920) also shows the "all or none" nature of nervous conduction. The well known fact that a strong stimulus applied to a nerve trunk causes an action current of greater amplitude, and a greater muscular response than a weaker stimulus appears to contradict the all or none principle. The nerve trunk, however, is composed of many fibers each of which supplies a group of muscle fibers. The weak stimulus excites only a proportion of the units of the nerve, a maximal stimulus excites them all. For example, the *cutaneous dorsæ muscle* of the frog is supplied by a nerve which contains only 8 or 9 fibers, each of these innervates about 20 muscle fibers. Keith Lucas found that when the nerve was stimulated by shocks, gradually increasing in intensity, the muscular responses did not show a similar continuous rise in amplitude, on the contrary, the responses of the muscle increased in a series of well-defined steps, that is, increasing the stimulus intensity produced no effect for a time upon the amplitude of the muscular response, but then a slight increase in strength of stimulus produced a sudden rise in amplitude. The steps were never greater in number than the number of fibers, and were due, it was concluded, to additional fibers

becoming excited as the strength of stimulus reached a certain value.

It must also be remembered that the all or none principle applies only for the condition of the nerve at the point where, and the moment when, the impulse arises. A stimulus which will give rise to a response of a certain magnitude under one condition of the nerve may give a much smaller response under other conditions, e.g., during the relative refractory period (p. 922), narcosis, oxygen lack, etc.

Variations in the frequency of the impulses in a single nerve fiber

The magnitude of the muscular response is determined not only by the number of neuromuscular units excited but also by the frequency of the impulses transmitted along the individual nerve fibers. When a motor nerve is stimulated maximally by a single induction shock, an impulse passes along each nerve fiber and the "volley" of impulses upon reaching the muscle causes a single contraction of all its fibers—a so-called muscle twitch. If the nerve is stimulated again after a certain brief interval a second volley of impulses is discharged and a second contraction occurs, which becomes superimposed upon the first to produce a greater muscular response. This is spoken of as *summation of contractions*. A series of volleys reaching the muscle at sufficiently short intervals will prevent any relaxation of the muscle between separate volleys. That is, the individual contractions become fused to produce a sustained maximal response or *tetanus*. Adrian and Bronk isolated a *single* fiber of the phrenic nerve in the cat and recorded the action currents passing along the fiber during normal respiration and after clamping the trachea. The action currents increased in frequency from between 20 and 30 per second during quiet breathing to between 50 and 80 during the forcible respiratory movements induced by the asphyxia. But no change in the amplitude or form of the electrical waves occurred. At the lower rate of discharge the individual contractions were incompletely fused (*incomplete tetanus*). At the higher rates the tetanus was complete and the tension developed by the contraction maximal. It might be thought that when the impulses are of low frequency and there is, in consequence, incomplete fusion of successive twitches the contractions of a muscle would be uneven and jerky. This would be so did the impulses discharged from the nerve centers travel synchronously along the

different fibers composing the motor nerve. At low frequency, however, the impulses are discharged asynchronously, i.e., not as a volley but rather like a scattered fire of rifle shots. Groups of muscle fibers are therefore activated asynchronously, the contractions of different groups overlap and a smooth steady contraction of the muscle as a mass results. At high frequencies the impulses are discharged along the separate nerve fibers synchronously, i.e., in a series of volleys, but owing to the fusion of the successive contractions the action of the muscle as a whole is also smooth and sustained. Results comparable to those described for the phrenic were obtained from a single fiber of a motor nerve supplying a skeletal muscle (nerve to the peroneus longus, to the tibialis anticus or to the quadriceps). A reflex discharge of impulses was produced by stimulation of the foot and the impulses recorded from the motor nerve, all the fibers of which except one had been severed. The impulse frequency varied from 20 to 90 per second with the strength of the stimulus. *The amplitude of the recorded electrical changes was not increased by the stronger stimulation*, i.e., the action potential is "all or none in nature."

By inserting a pair of very closely approximated electrodes into a muscle, it was found possible to record the action currents of a small group of muscle cells supplied by a single nerve fiber (Adrian and Bronk). A fine copper wire (gauge 36), enamelled except for its tip, was passed down a small hypodermic needle. The exposed wire projecting from the point of the needle forms one electrode and is connected to the input of an amplifier. The needle itself constitutes the other electrode and is connected to earth. The action currents may be led, after amplification, to a capillary electrometer or oscillograph and recorded photographically, or converted into sound by a loud speaker, and their rhythm picked up by ear. When the concentric needle electrodes were inserted into the human triceps and a voluntary movement made, the records indicated that the impulses arriving along a single nerve fiber varied with the strength of the contraction from 5 to 50 per second. No electrical changes were observed when the muscle was completely relaxed. At the higher rate the record showed minor waves. These were interpreted as indicating that additional neuromuscular units also came into action during the stronger contraction. At the lower frequency the active fiber groups are fewer and consequently more widely separated, the electrical responses from neighboring groups therefore do not complicate the electrical record of the group immediately surrounding the electrodes. When certain muscles were excited reflexly, e.g., peroneus longus, or tibialis

anticus of the cat, there was little evidence that the stronger contraction was due to more fiber groups coming into play, i.e., to impulse discharges over a greater number of motor nerve fibers, the grading of the contraction appeared to be due chiefly to variations in the frequency of the impulses.

The absolute and relative refractory periods of nerve

Within a certain brief interval following the passage of an impulse along the nerve fiber a second stimulus, however strong, is unable to evoke a response. This interval is called the absolute refractory period. In a frog's sciatic nerve at a temperature of about 15°C the absolute refractory period has a duration of between 2 and 3 msec. Its duration is the same or slightly longer than that of the action potential "spike" (see pp 923-925). It is much shorter in mammalian nerve (about 0.4 msec). This period during which the nerve is absolutely refractory is succeeded by one in which the nerve, though it will not respond to as weak a

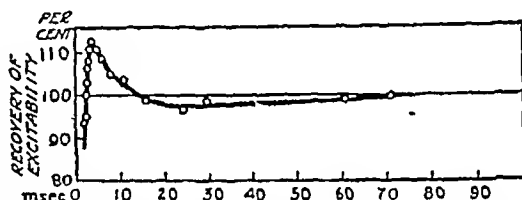


FIG 62.11 Recovery curve of nerve. Cat's saphenous nerve stimulated at O. (From Gasser and Grundfest)

stimulus as it did before the passage of the impulse, will respond to a somewhat stronger one. The excitability of the nerve gradually increases and the strength of stimulus necessary for excitation becomes progressively less (fig 62.11). In the end, the restoration of excitability is complete and the nerve responds to a stimulus of no greater strength than that which is capable of exciting a resting nerve. This period following the absolute refractory phase and during which the excitability gradually rises to normal is called the *relative refractory period*. It lasts for from 10 to 20 msec or at any rate the excitability of the nerve has returned to about 95 per cent of the resting value by this time. (Full recovery however may not be attained until the lapse of 100 msec.) It should be pointed out that the failure of the nerve to conduct a second impulse is not due simply to lowered excitability at the point in the nerve where the original stimulus was applied, for during the absolute refractory period a stimulus applied to

any other point upon the nerve likewise fails to set up an impulse. The passage of the impulse along the nerve leaves in its wake a change of state like a trail of ash after the ignition of a powder fuse. For the moment the impulse consumes the entire resources of the nerve fiber (Adrian). The burned fuse must have its store of energy replenished by laying a fresh train of powder grains before a second spark can traverse the path of the first. So also, a certain time is required for the changes associated with the passage of the impulse to become reversed and the nerve restored to its resting condition (polarized state, see also p 926).

The refractory period renders a continuous excitatory state of the nerve impossible just as the corresponding period in cardiac muscle assures rhythmical contractions and prevents summation and tetanus. Fusion or summation of impulses does not occur. The refractory period obviously must also limit the frequency of the impulses. In the mammal the absolute refractory period is about $\frac{1}{1000}$ second. The intervals between impulses cannot be shorter than the absolute refractory period, the maximum impulse frequency is therefore around 1000 per second. At this rate the impulses are travelling in the *relative* refractory period of their predecessors and are consequently weaker. In frog nerve with its refractory period of from 2 to 3 milliseconds, the maximal impulse frequency is between 250 and 300 per second.

THE SUPERNORMAL AND SUBNORMAL PHASES The relative refractory period is followed by one lasting for from 3 to 15 milliseconds, in which the nerve fiber is hyperexcitable. This is followed in turn by a state of subnormal excitability which persists for from 15 to 70 m.sec.

THE ELECTRICAL CHANGES IN NERVE

The current of injury

When a pair of electrodes are placed a short distance apart upon the surface of an uninjured and resting nerve (or muscle) and connected through a galvanometer, no current flows and no deflection of the instrument occurs, since the entire surface of the nerve (or muscle), and, therefore, the tissue beneath each electrode, is of the same electrical potential. When, however, one part of the tissue is injured the membrane at this point becomes depolarized (see membrane theory, p 926) and, in consequence, negative to the positively

charged uninjured surface. When the two electrodes are now placed one on the injured and the other on the uninjured part and connected through a galvanometer a constant current flows through the instrument from the uninjured (+) to the injured (−) section and in the opposite direction through the length of the tissue. This is known as the *current of injury* (or *demarcation current*) (fig 62 12, I).

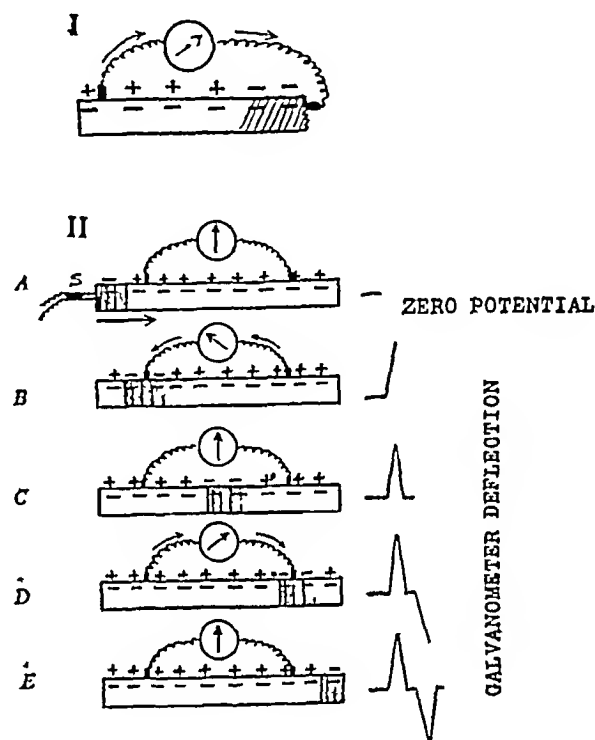


FIG 62 12 I, current of injury, II, current of action. Description in text. The small arrows indicate the flow of the current in the galvanometer leads, the large arrow, in II A, indicates the direction of spread of the wave of contraction which is represented by the stippled area.

The current of action

Active tissue is also relatively negative to resting tissue. The former corresponds to the zinc or negative pole of a battery, the latter to the copper or positive pole. If, therefore, two electrodes are placed upon a section of nerve and connected through a galvanometer, as shown in fig 62 12 II, A, stimulation of the nerve at S causes a movement of the galvanometer indicator first in one direction and then in the other.³ The movement may be photographed, when a diphasic curve—a wave above the base line followed by one in the

³ Today a cathode ray oscillograph is employed (p 924) which, being of extreme sensitivity, enables very weak currents to be picked up and recorded.

opposite direction—is obtained. This is the result of the propagation of the wave of excitation along the nerve from one electrode to the other. When the wave reaches the tissue beneath the first electrode this region becomes negative to the resting tissue beneath the second electrode. A current flows through the galvanometer from the resting to the active region. A wave above the base line is inscribed in the record (fig. 62 12 II, B). During the lapse of time from the passage of the wave from beneath the first electrode to its arrival at the second no current flows, and the galvanometer indicator returns to the zero position (fig. 62 12 II, C). The arrival of the excitation wave beneath the second electrode now renders the tissue here negative to the resting tissue under the first electrode, and a current flows through the galvanometer, but in a direction opposite to that of the current previously set up. As the nerve beneath the second electrode returns to the inactive state the galvanometer comes to rest. A wave below the base line is thus recorded (fig. 62 12 II, D and E). The diphasic electrical change is called the *current of action*. If a current of injury is set up by crushing the nerve beneath the second electrode, an action current initiated by stimulating the nerve on the near side of the first electrode, since it will be opposite in direction to the current of injury, will cause a reduction in the strength of the latter, as indicated by a movement of the galvanometer needle. The reduction thus caused in the strength of the current of injury is spoken of as the *negative variation*. The excitation wave does not reach the tissue beneath the second electrode since it is blocked by the injured tissue so a second oppositely directed wave does not appear, i.e., the variation is *monophasic*. The injury potential can be compensated for by balancing against it an opposing current, a record of a monophasic action potential—the “spike potential”—is thus obtained. The term “negative” is used in the sense that active tissue under the first electrode becomes *less positive* than during the resting state.

So far as is known the propagated disturbance in nerve is invariably accompanied by an electrical change. In fact there is no evidence to indicate that the nerve impulse and the traveling electrical change are not one and the same. At any rate they are very intimately related and the investigation of the electrical changes in the nerve fibers during excitation is therefore the most convenient and at the same time the most sensitive and

reliable means we possess for studying the frequency, speed and strength of the nerve impulse. For example, the number of successive electrical changes traversing the nerve is taken as representing the frequency of the impulses, the speed and strength of the impulse is judged from the speed and amplitude of the electrical response. The electrical waves are of the same general form in various types of nerve fiber, this is taken as an indication that the impulses in the different nerve fibers are fundamentally similar.

After-potentials Following the main action potential or “*spike*” potential as it is now usually called, a series of smaller electrical changes appear in tracings obtained by the more sensitive methods of recording. These are known as *after potentials*. They appear following a series of spikes, but in their simplest form, after a single response. They consist of an initial negative potential (i.e. of the same direction as the spike) followed by a positive potential of much smaller magnitude (about 0.2 per cent of the height of the spike), but of longer duration. The negative after potential has a duration in rapidly conducting fibers of about 15 milliseconds, the positive after potential of 70 milliseconds or more. The after potentials show much greater variability with experimental conditions than does the spike. A tetanizing current increases both negative and positive after-potentials, with such a current a second positive potential frequently appears. The spike coincides approximately with the absolute refractory period, the negative and positive after potentials with the super-normal and subnormal phases of excitability, respectively.

The “spike” and the after-potentials can also be recorded from the spinal cord during stimulation of a dorsal nerve root. The spike potential arises from the continuation into the cord of the fibers of the dorsal roots, but Gasser concludes for several reasons that the after potentials are derived from internuncial neurons.

The compound nature of the action current recorded from a nerve trunk

Erlanger, Bishop and Gasser studied the action potential of mixed nerve trunks by means of the cathode ray oscillograph.

Unlike any instrument previously employed for this purpose, e.g., the string galvanometer or the capillary electrometer, the moving part of this instrument—a stream of electrons—possesses practically no mass and is in consequence inertialess. It is therefore capable of recording almost infinitely rapid changes in electrical potential. The instrument consists of an evacuated tube through which the electron stream is thrown against a fluorescent screen upon which it produces a

spot of light. On either side of the electron stream is placed a vertical plate. A potential difference is created between the pair of plates, the electric field thereby set up across the path of the stream deflects it horizontally, and sweeps it across the screen. The spot of light is thereby converted into a horizontal streak. By means of a rotating commutator the deflections are repeated from 10 to 20 times per second. A second pair of horizontal plates is placed one above, the other below, the electron stream. These are connected with the nerve whose action current is timed to reach them at the instant that the stream is deflected horizontally by the vertical plates. A vertical deflection of the electron stream results with the production of a standing wave which is photographed and a permanent record thus obtained. The speed of the horizontal movement of the spot of light enables the time factor to be calculated, and can be varied by altering the potential applied to the vertical pair of plates, the horizontal movement corresponds to the movement of a kymograph, though of course its rate is very many times faster. The upward deflection is analogous to the rise of a muscle lever. The magnitude of the action potential is determined from the height of the wave. Before reaching the recording system the action current is amplified several thousand times by passing it through a three-stage amplifier.

Upon analysis of the electrical potentials of mammalian nerves, Erlanger and Gasser showed that the "spike" is actually compound and represents the fusion of the potentials of three main types of nerve fiber, which are referred to as the A, B, and C groups. Several properties of nerve are correlated with the diameters of the fibers, the larger the fiber diameter, the higher is the conduction velocity, the lower the threshold of excitation, the greater the magnitude of electrical response, but the shorter its duration, and the shorter the refractory period, latent addition⁴ and chronaxie. The relationship of conduction velocity to outside diameter of the nerve fiber is a linear one (fig 62 13). The amplitude of the recorded potential is also linearly related to the fiber diameter.

The A group is composed of the largest fibers, 1 to 20 micra in diameter, with conduction rates from 5 meters per second or less for the smallest fiber to 100 meters per second for the largest. This group has been analyzed into four subsidiary groups labelled α , β , γ , δ in accordance with and in this order of fiber diameter (fig 62 14). The A groups are all myelinated, are both sensory and motor and are found in such somatic nerves as the

⁴ This is the period during which a subliminal stimulus is capable of summing with a previous one and of raising excitation to threshold value.

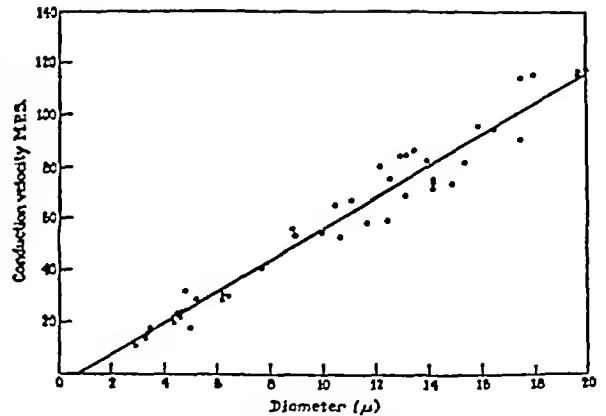


FIG 62 13 Linear relation between diameter and conduction velocity of mammalian nerve fibers. Each point represents a determination of the maximum conduction velocity in meters per sec. and of the diameter in micra of the largest fiber of an individual nerve. Dots = adult nerves. Circles = immature nerves. (After Hursh)

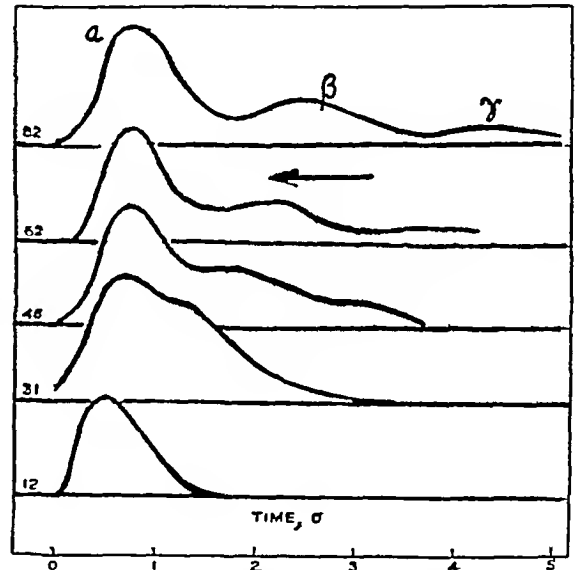


FIG 62 14 Cathode ray oscillograph records of the action currents in the sciatic nerve of the bullfrog after conduction from the point of stimulation through the distances (in millimeters) shown at the left. The action potentials might be compared to runners in a race who become separated along the course as the faster contestants outstrip the slower, thus in a record at 82 mm from the point of stimulation three waves are shown, whereas at 12 mm the potentials are fused, and only one large wave appears. The delta wave is not shown in the record. (Modified from Erlanger, Bishop and Gasser)

sciatic and saphenous nerves. The alpha fibers of this group are absent from cutaneous nerves.

Less study has been made of the B and C groups. The B fibers are myelinated and have diameters from 1 to 3 micra, and conduction velocities from about 3 meters per second to 14 meters

per second. The B fibers are found solely in autonomic (preganglionic and in some myelinated postganglionic) nerves. The C group, composed of the smallest fibers (less than 1 micron in diameter), are unmyelinated and have a conduction rate of around 2 meters per second or less. They transmit pain impulses and possibly those of pressure. The A group of fibers make by far the greatest contribution to the compound spike potential, and the C group the least. The electrical potentials recorded from both A and C fibers show both negative and positive after potentials, but the B group shows no negative after potential with a single response, though it appears upon repetitive stimulation. The B fibers are the most susceptible to asphyxia, the C fibers the least so.

The linear relationship between fiber diameter and conduction velocity holds also for growing nerves of young animals (Hursh). During growth the diameters of the nerve fibers in the nerve of the leg, for example, increase as the nerve lengthens. Conduction velocity increases proportionately so that the time taken for an impulse to travel from the toes of a kitten a few days old to the spinal cord is the same as for a full grown cat. Thus the kitten and the cat react to stimulation with about equal promptness.

The diameters of the regenerating fibers in a sectioned or crushed nerve also enlarge gradually and conduction velocities increase accordingly (Berry and associates), the relationship again being a linear one. The maximum conduction velocity is not reached until maximum diameter of the fiber is attained. If the axons of the nerve alone are interrupted, the sheaths of the nerve fibers remaining intact, the diameters and conduction velocities may reach those of the normal nerve, this rarely occurs if the nerve has been completely severed.

THE MEMBRANE THEORY (BERNSTEIN) OF NERVOUS CONDUCTION

Conduction, according to this theory, is a surface phenomenon. The nerve fiber is surrounded by a semi permeable membrane or surface film which is polarized when the nerve is at rest. That is, the surface film separates a layer of cations on its outer side from a layer of anions on its inner side. A stimulus applied to the nerve increases the permeability of the membrane at the point of stimulation with the result that a redistribution of ions and depolarization of the membrane occur. This point of the nerve becomes thereby relatively

negative to the inactive (polarized) section of nerve immediately adjacent. A potential difference is set up and a current flows between the active and inactive sections. This secondary current in turn causes depolarization and activation of the adjoining region which, being now relatively negative to the next section, results in a current being again set up between these two which depolarizes the latter. The currents set up between contiguous, inactive and active regions serve as successive stimuli and the wave of depolarization spreads down the nerve, the disturbance set up by the original artificial stimulus is in this way propagated automatically (see fig. 62.15). Thus, nerve impulses are "transient self-mending electrical leaks traveling along nerve fibers" (Sherrington). The depolarized state persists for a short time after the

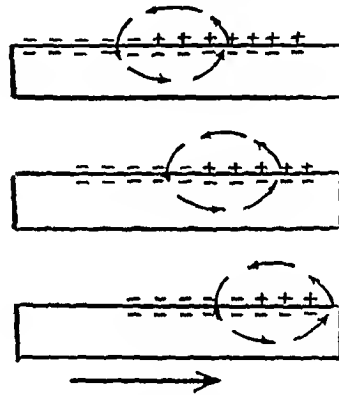


FIG. 62.15 Illustrating nervous conduction according to the membrane theory (see text)

passage of the impulse, during this time the nerve is refractory. The return of excitability is dependent upon the restoration of the polarized state. According to the membrane theory the nerve impulse is simply the propagated wave of depolarization.

Perhaps the greatest support for the membrane theory is afforded by the experiments of Lillie who has prepared a metal model which behaves in a manner comparable to that of nerve. A film of oxide forms upon an iron wire placed in strong nitric acid. When such an oxide-coated wire is then immersed in a weak solution of acid, which would cause gradual solution of an untreated wire, no reaction results. The metal is in a so-called passive state, being comparable to a resting nerve. When, however, the wire is "stimulated", e.g., touched at one point with active iron or some other base metal, scratched with a piece of glass

in order to break the protective oxide film, an electric current applied to it, a reaction (electrochemical reduction) is set up which is accompanied by effervescence and the formation of a dark colored layer of oxide, proceeds down the wire. When two parts of the wire are connected with a galvanometer a current flows through the circuit causing the spread of the reaction. If the salt bathing the wire is of suitable concentration the protective film reforms in the wake of the excitation. If a small current is applied after the film has reformd the effects are repeatable of the phenomenon. The rate of rate of the reaction of the wire is to the nerve as by analogy. If there are electrical reactions, the flow of electrons is analogous to that (see the potential difference). Suppose the potential were an electrochemical difference existing between two wires, then a cell and the current of ions. Differences at a point in the filament cause the region to become sensitive to a local reaction as a wave of depolarization is propagated to the rest of the wire.

In a model of the nerve model is the following example the nerve is as to that of nerve.

(1) The rate of propagation of the reaction is of the order of the nerve.

(2) The rate of reaction of a certain rate is (1) the rate of reaction of the nerve.

(3) The rate of reaction of the nerve is (1) the rate of reaction of the nerve.

(4) When a reaction is completed to activate the nerve it is (1) the rate of reaction of the nerve.

(5) The rate of reaction of the nerve is (1) the rate of reaction of the nerve.

In the wire model a reaction is able to propagate in the direction of the oxide deposit is produced. It is possible that the excitation of nerve is associated with a corresponding change (increase in permeability) in the surface film surrounding the nerve fiber. Though such a permeability change has been demonstrated in certain slowly conducting protoplasmic systems, the evidence for its occurrence in nerve is indirect. In the case of nerve and other protoplasmic systems the alteration in permeability is associated with chemical changes—the processes underlying the excitation of nerve and the transmission of the impulse are therefore, as in the case of the iron wire model, essentially electrochemical in nature.

METABOLISM OF NERVE AND BRAIN

It has been thought until comparatively recent years that a negligible expenditure of energy ac-

companies the transmission of the nerve impulse, the latter was believed to be dependent upon physical rather than upon chemical processes. Though it was recognized that conduction eventually failed in the absence of oxygen this fact did not prove that conduction itself was due to oxidative processes. In order to show this it would be necessary to demonstrate an increase in the oxygen consumption and in the carbon dioxide production during activity. An increased carbon dioxide production of active nerve over that of a resting nerve was observed by Iashiro in 1913. This was confirmed by Parler, and an increased oxygen consumption during excitation was shown by Tenn. Further attempts (Helmholtz, 1848) to demonstrate heat production in active nerve were unsuccessful, but in 1926 Hill, using a thermopile composed of over two hundred thermocouples, detected a rise of 10^{-6} °C in the temperature of a nerve as a result of stimulation lasting 10 seconds.

HEAT PRODUCTION IN NERVE

The resting sciatic nerve of the frog in oxygen at 20°C. was found by Bergsma to generate 4.14 calories per gram of nerve per minute. In nitrogen the resting heat production falls gradually reaching, at the end of about 3 hours, to 20 or 25 per cent of the value in oxygen and remains steady at this low level. Upon the readmission of oxygen to the asphyxiated nerve, heat production rises rapidly, extra heat being produced over that produced previously by the nerve in oxygen. This extra heat amounts to some 15 or 20 per cent of the "muscle" heat in nitrogen (i.e., of the amount equal to the difference between heat production in oxygen and in nitrogen). The heat generated in nitrogen is attributed mainly to an oxidative process dependent upon a store of oxygen in the nerve tissue, and to the breakdown of carbohydrate to lactic acid, and possibly also to phosphocreatine breakdown. In the steady state the heat production is attributed to lactic acid production from carbohydrate. The extra heat produced upon the readmission of oxygen is probably in part waste heat in the process of phosphocreatine resynthesis, and possibly also to the oxidation of some of the lactic acid accumulated during the anaerobic period.

During stimulation of the nerve in oxygen, at the rate of 250 shocks per second, an increase in heat production occurs of 10×10^{-6} calories per gram

* Small calories

of nerve per second. At this rate of stimulation the heat resulting from a single impulse is about 1.4×10^{-7} calories per gram of nerve fiber. Though the heat production per gram of resting nerve is nearly as great as that of resting muscle, during activity the heat produced by nerve is only $\frac{1}{400,000}$ of that generated by muscle of equal weight and stimulated to the same degree.

As in the case of muscle, the heat produced during activity is given off in two batches—the *initial heat* and the *delayed* or *recovery heat*. The initial heat is less than 4 per cent of the total heat,⁶ the ratio of initial to delayed heat being 1:30. The rate of generation of the initial heat is some 5000 times greater than that of the

delayed heat.⁷ Increase in the *strength* of the stimulus does not increase the heat production, whereas increase in the *frequency* of stimulation causes an increase in heat production up to 25 per cent. The heat per second, however, does not increase proportionately with the increase in frequency of stimulation, so the heat per impulse is actually reduced. When the shocks are at the rate of 280 per second the heat production per second is maximal and the heat per impulse (since each impulse is travelling in the relative refractory period of its predecessor) is only a quarter of that generated by a single isolated impulse. The heat production of the central nervous system is enormously greater than that of nerve fibers amounting to from 600 to over 2000×10^{-6} per gram per second for the spinal cord of the frog.

CARBON DIOXIDE PRODUCTION AND OXYGEN CONSUMPTION

The resting sciatic of the frog produces in the neighborhood of 0.6 cu. mm. CO_2 per gram of nerve per minute. The corresponding O_2 consumption is about 0.7 cu. mm. The resting respiratory quotient is therefore around 0.8. During activity an extra 0.25 cu. mm. of O_2 per gram of nerve per minute is consumed and a somewhat smaller quantity of extra CO_2 produced. The extra metabolism of the nerve resulting from activity has an R.Q. of about 0.90 (Meyerhof and Schmidt).

The extra oxygen consumption resulting from nerve activity occurs during the period of delayed heat. It continues for a considerable time—15 minutes or more after the impulse has passed. The quantity of oxygen consumed agrees well with the heat produced at this time, upon the basis that the latter is the result of the oxidation of ordinary food materials. Like the heat production the oxygen consumption per impulse falls with a rise in frequency of stimulation, though of course the total consumption per minute increases. Increasing the strength of the stimulus beyond that necessary to excite all the fibers does not increase the oxygen consumption.

CHEMICAL CHANGES IN NERVE

(a) IN THE ABSENCE OF OXYGEN. Placed in nitrogen a *resting* nerve undergoes a reduction in

⁷ The cause of the initial heat is unknown. It may be chemical in nature and due to the breakdown of phosphocreatinine or may be derived, as Hill suggests, from an electrical source, namely, the discharge of an electric double layer—a condenser—located at the surface of the fiber. See A. V. Hill, "Chemical wave transmission in nerve." Cambridge, University Press, 1932.

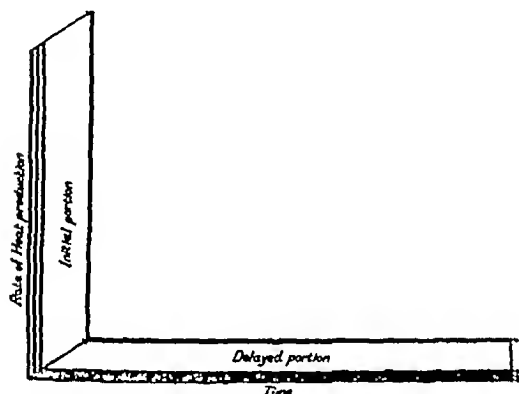


FIG. 62.16. Diagrammatic representation (not to scale) of the heat production due to each nervous impulse, indicating how the observed heat resulting from a tetanic stimulation is built up of these units. The horizontal lines represent the delayed heat, starting at a maximum rate of 0.05 "C" units and slowly falling to zero in about 10 mins. The vertical lines represent the initial heat which probably is largely produced during less than 4 milliseconds at a rate 5000 times greater than that at the start of the delayed phase (after Gerard).

delayed heat (fig. 62.16). The former coincides with or follows immediately upon the "spike" potential being an intense explosive outburst which, with a single stimulus, lasts for only a few milliseconds. The delayed or recovery heat which is given out after the period of activity declines rapidly at first and then more slowly.

The delayed heat is accompanied by the consumption of oxygen and is evolved in two stages. The first of these lasts for a few seconds, and the quantity of heat is of small magnitude, the second lasts for from 10 to 30 minutes and contributes the greater proportion of the

⁶ Hill found that the initial heat of crab nerve was only 2.25 per cent of the total heat, which, however, was much greater (2.5×10^{-7} per second per gram of fiber) than that of frog nerve. Beresina and Feng obtained similar values for crab nerve.

its glycogen and free sugar contents and an accumulation of lactic acid. The lactic acid production proceeds slowly and in about 3 hours has reached a maximum rate of 7 mg per 100 grams of nerve per hour, it then falls gradually to zero. The total acid production is about 100 mg per 100 grams of nerve (0.1 per cent). This is not attained until the end of 24 hours, at which time the carbohydrate store of the nerve becomes exhausted. Immersion of the nerve in a solution of glucose (but not of galactose or fructose) causes lactic acid production to continue at the maximum rate (7 mg per cent) for days—or until conversion of the added glucose has occurred. When the nerve is *stimulated* in nitrogen there does not result, however, as in the case of muscle deprived of oxygen, an increased accumulation of lactic acid. Also when oxygen is re-admitted lactic acid does not disappear or does so very slowly. Oxygen consumption of the nerve is somewhat greater than usual after a period in nitrogen—an indication that nerve runs into debt for oxygen during a period of anoxia. In the absence of oxygen, nerve, unlike muscle, continues to respond to stimulation for a considerable time. It shows a progressive fall in excitability but does not fail to conduct until the lapse of about 3 hours. The action current or the current of injury is gradually reduced during this time but rises again when oxygen is re-admitted. A fall in phosphocreatine content and a rise in inorganic phosphorus also occur.

In the asphyxiated nerve the delayed heat production declines no more rapidly (if anything less so) than the initial heat, whereas in the case of a muscle contracting in the absence of oxygen most of the recovery heat is abolished (p. 724). In nerve, therefore, *both conduction and the recovery process can apparently be accomplished for a time after the external oxygen supply has been cut off*.

(b) IN THE PRESENCE OF OXYGEN. In the case of the excised *resting nerve* free sugar gradually disappears but the glycogen content remains unchanged and *no lactic acid is formed*. These observations indicate that the function of the resting nerve is probably maintained in part by energy derived from the combustion of sugar. Energy is also probably derived from the breakdown of phospholipids. *During activity* phosphocreatine is broken down but the glycogen content of the nerve remains unaltered. Sugar does not disappear more rapidly than during rest, nor is lactic acid produced. The glycogen-lactic acid cycle which is so promi-

nent in the metabolism of muscle therefore appears to play no part in nerve fiber conduction. The immediate source of the energy for conduction is believed to be derived from the breakdown of phosphocreatine which is resynthesized after the passage of the impulse. The ultimate source of the energy for the recovery process in nerve is unknown.

It has been thought that the nerve fiber was incapable of oxidizing lactic acid or of synthesizing lactic acid to glycogen, but, as mentioned above, when an accumulation of lactic acid occurs as a result of asphyxia small amounts do disappear upon the re-admission of oxygen. Also, in a nerve soaked in sodium iodoacetate the phosphocreatine disappears and the nerve soon fails to respond, but if lactate be supplied the survival time is considerably lengthened and oxygen consumption increased (Feng). (The addition of lactate causes no increased oxygen consumption by normal nerve.) These and other observations indicate that the inability of the nerve to oxidize lactic acid is not absolute but that, under certain circumstances at any rate, such does occur.

Small quantities of *ammonia* (about 0.3 mg per cent per hour) are produced by resting nerve and about double this amount during activity. The source of the ammonia is probably adenylypyrophosphate. The *inorganic phosphate* also increases during activity as a result of the breakdown of phosphocreatine and adenylypyrophosphate and, possibly, of phospholipids.

Potassium of nerve. The nerve fiber is exceptionally rich in potassium. Cowan found that the potassium concentration in the nerve of the crab is some 13 times greater than that in crab's blood. Ratios as high as 65:1 for the potassium concentrations on the inner and outer sides of the membrane have been reported for vertebrate nerve. It is upon this difference in the potassium concentrations that the resting potential (polarization) of the nerve is believed, according to the Bernstein hypothesis, to depend. But when the nerve is stimulated or deprived of oxygen, potassium diffuses rapidly into the surrounding fluid, but is restored again (repolarized) during rest or, in the case of the asphyxiated nerve, after the readmission of oxygen. A nerve at rest and adequately supplied with oxygen does not lose potassium and a potential difference between the surface of the fiber and its interior is maintained. Increasing the concentration of potassium in the fluid bathing the nerve will tend to reduce the

Smith

of nerve per second. At this rate of stimulation the heat resulting from a single impulse is about 1.4×10^{-7} calories per gram of nerve fiber. Though the heat production per gram of resting nerve is nearly as great as that of resting muscle, during activity the heat produced by nerve is only $\frac{1}{400,000}$ of that generated by muscle of equal weight and stimulated to the same degree.

As in the case of muscle, the heat produced during activity is given off in two batches—the *initial heat* and the *delayed or recovery heat*. The initial heat is less than 4 per cent of the total heat,⁶ the ratio of initial to delayed heat being 1:30. The rate of generation of the initial heat is some 5000 times greater than that of the

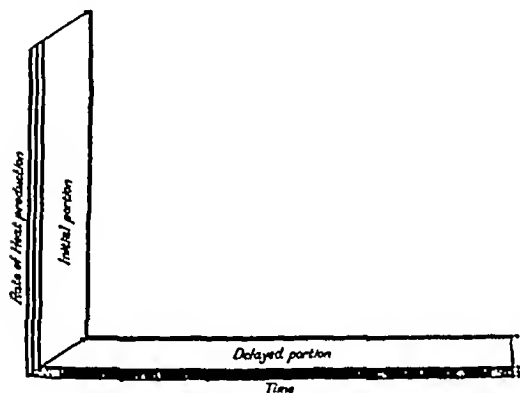


FIG. 62.16 Diagrammatic representation (not to scale) of the heat production due to each nervous impulse, indicating how the observed heat resulting from a tetanic stimulation is built up of these units. The horizontal lines represent the delayed heat, starting at a maximum rate of 0.05 "C" units and slowly falling to zero in about 10 mins. The vertical lines represent the initial heat which probably is largely produced during less than 4 milliseconds at a rate 5000 times greater than that at the start of the delayed phase (after Gerard).

delayed heat (fig. 62.16). The former coincides with or follows immediately upon the "spike" potential being an intense explosive outburst which, with a single stimulus, lasts for only a few milliseconds. The delayed or recovery heat which is given out after the period of activity declines rapidly at first and then more slowly.

The delayed heat is accompanied by the consumption of oxygen and is evolved in two stages. The first of these lasts for a few seconds, and the quantity of heat is of small magnitude, the second lasts for from 10 to 30 minutes and contributes the greater proportion of the

⁶ Hill found that the initial heat of crab nerve was only 2.25 per cent of the total heat, which, however, was much greater (2.5×10^{-2} per second per gram of fiber) than that of frog nerve. Beresina and Feng obtained similar values for crab nerve.

delayed heat.⁷ Increase in the *strength* of the stimulus does not increase the heat production, whereas increase in the *frequency* of stimulation causes an increase in heat production up to 25 per cent. The heat per second, however, does not increase proportionately with the increase in frequency of stimulation, so the heat per impulse is actually reduced. When the shocks are at the rate of 280 per second the heat production per second is maximal and the heat per impulse (since each impulse is travelling in the relative refractory period of its predecessor) is only a quarter of that generated by a single isolated impulse. The heat production of the central nervous system is enormously greater than that of nerve fibers amounting to from 600 to over 2000×10^{-4} per gram per second for the spinal cord of the frog.

CARBON DIOXIDE PRODUCTION AND OXYGEN CONSUMPTION

The resting sciatic of the frog produces in the neighborhood of 0.6 cu. mm. CO_2 per gram of nerve per minute. The corresponding O_2 consumption is about 0.7 cu. mm. The resting respiratory quotient is therefore around 0.8. During activity an extra 0.25 cu. mm. of O_2 per gram of nerve per minute is consumed and a somewhat smaller quantity of extra CO_2 produced. The extra metabolism of the nerve resulting from activity has an R.Q. of about 0.90 (Meyerhof and Schmidt).

The extra oxygen consumption resulting from nerve activity occurs during the period of delayed heat. It continues for a considerable time—15 minutes or more after the impulse has passed. The quantity of oxygen consumed agrees well with the heat produced at this time, upon the basis that the latter is the result of the oxidation of ordinary food materials. Like the heat production the oxygen consumption per impulse falls with a rise in frequency of stimulation, though of course the total consumption per minute increases. Increasing the strength of the stimulus beyond that necessary to excite all the fibers does not increase the oxygen consumption.

CHEMICAL CHANGES IN NERVE

(a) IN THE ABSENCE OF OXYGEN. Placed in nitrogen a *resting* nerve undergoes a reduction in

⁷ The cause of the initial heat is unknown. It may be chemical in nature and due to the breakdown of phosphocreatinine or may be derived, as Hill suggests, from an electrical source, namely, the discharge of an electric double layer—a condenser—located at the surface of the fiber. See A. V. Hill, "Chemical wave transmission in nerve." Cambridge, University Press, 1932.

its glycogen and free sugar contents and an accumulation of lactic acid. The lactic acid production proceeds slowly and in about 3 hours has reached a maximum rate of 7 mg per 100 grams of nerve per hour, it then falls gradually to zero. The total acid production is about 100 mg per 100 grams of nerve (0.1 per cent). This is not attained until the end of 24 hours, at which time the carbohydrate store of the nerve becomes exhausted. Immersion of the nerve in a solution of glucose (but not of galactose or fructose) causes lactic acid production to continue at the maximum rate (7 mg per cent) for days—or until conversion of the added glucose has occurred. When the nerve is *stimulated* in nitrogen there does not result, however, as in the case of muscle deprived of oxygen, an increased accumulation of lactic acid. Also when oxygen is re-admitted lactic acid does not disappear or does so very slowly. Oxygen consumption of the nerve is somewhat greater than usual after a period in nitrogen—an indication that nerve runs into debt for oxygen during a period of anoxia. In the absence of oxygen, nerve, unlike muscle, continues to respond to stimulation for a considerable time. It shows a progressive fall in excitability but does not fail to conduct until the lapse of about 3 hours. The action current or the current of injury is gradually reduced during this time but rises again when oxygen is re-admitted. A fall in phosphocreatine content and a rise in inorganic phosphorus also occur.

In the asphyxiated nerve the delayed heat production declines no more rapidly (if anything less so) than the initial heat, whereas in the case of a muscle contracting in the absence of oxygen most of the recovery heat is abolished (p. 724). In nerve, therefore, *both conduction and the recovery process can apparently be accomplished for a time after the external oxygen supply has been cut off*.

(b) IN THE PRESENCE OF OXYGEN. In the case of the excised *resting nerve* free sugar gradually disappears but the glycogen content remains unchanged and *no lactic acid is formed*. These observations indicate that the function of the resting nerve is probably maintained in part by energy derived from the combustion of sugar. Energy is also probably derived from the breakdown of phospholipins. *During activity* phosphocreatine is broken down but the glycogen content of the nerve remains unaltered. Sugar does not disappear more rapidly than during rest, nor is lactic acid produced. The glycogen-lactic acid cycle which is so promi-

nent in the metabolism of muscle therefore appears to play no part in nerve fiber conduction. The immediate source of the energy for conduction is believed to be derived from the breakdown of phosphocreatine which is resynthesized after the passage of the impulse. The ultimate source of the energy for the recovery process in nerve is unknown.

It has been thought that the nerve fiber was incapable of oxidizing lactic acid or of synthesizing lactic acid to glycogen, but, as mentioned above, when an accumulation of lactic acid occurs as a result of asphyxia small amounts do disappear upon the re-admission of oxygen. Also, in a nerve soaked in sodium iodoacetate the phosphocreatine disappears and the nerve soon fails to respond, but if lactate be supplied the survival time is considerably lengthened and oxygen consumption increased (Feng). (The addition of lactate causes no increased oxygen consumption by normal nerve.) These and other observations indicate that the inability of the nerve to oxidize lactic acid is not absolute but that, under certain circumstances at any rate, such does occur.

Small quantities of *ammonia* (about 0.3 mg per cent per hour) are produced by resting nerve and about double this amount during activity. The source of the ammonia is probably adenylypyrophosphate. The *inorganic phosphate* also increases during activity as a result of the breakdown of phosphocreatine and adenylypyrophosphate and, possibly, of phospholipins.

Potassium of nerve. The nerve fiber is exceptionally rich in potassium. Cowan found that the potassium concentration in the nerve of the crab is some 13 times greater than that in crab's blood. Ratios as high as 65:1 for the potassium concentrations on the inner and outer sides of the membrane have been reported for vertebrate nerve. It is upon this difference in the potassium concentrations that the resting potential (polarization) of the nerve is believed, according to the Bernstein hypothesis, to depend. But when the nerve is stimulated or deprived of oxygen, potassium diffuses rapidly into the surrounding fluid, but is restored again (repolarized) during rest or, in the case of the asphyxiated nerve, after the readmission of oxygen. A nerve at rest and adequately supplied with oxygen does not lose potassium and a potential difference between the surface of the fiber and its interior is maintained. Increasing the concentration of potassium in the fluid bathing the nerve will tend to reduce the

potential difference, this procedure also reduces the excitability of the nerve and with high concentrations excitability is completely lost but is restored again when the nerve is placed in sea water. The amplitude of the action current or of the current of injury is also markedly reduced by raising the potassium concentration on the outer side of the nerve. Furthermore, when the nerve is exposed to cold, "blocking" of the impulse occurs at a higher temperature than usual if the nerve has been soaked in a solution containing a high percentage of potassium.

The metabolism of the resting nerve may be considered in terms of the membrane theory (p. 926) to be directed toward securing a certain degree of impermeability of the surface film for the maintenance of the polarized state. The chemical changes associated with activity result in increased permeability and depolarization. During recovery the membrane is repolarized, the energy being furnished by oxidative processes—the battery is recharged. The source of the energy which enables a nerve to continue to conduct for such a long time in an atmosphere of nitrogen is a matter for speculation.

The *brain* (gray matter) when supplied with oxygen oxidizes glucose and, unlike the nerve fiber under the same conditions, produces lactic acid, though in small amounts. Lactic acid is also produced slowly when glucose is added to brain slices *in vitro* with free access to oxygen. Glucose is present in brain tissue in about the same concentration as in blood, and is the main substrate for the respiration of the gray matter. The respiratory quotient of the brain is around unity. As compared with muscle the gray matter contains only small amounts of glycogen. When the latter is added to brain slices it is broken down very slowly to lactic acid. Hexosephosphate, according to Ashford and Holmes, is not an intermediary in the oxidation of glucose by brain tissue, but the production of pyruvic acid appears to be an essential step. The utilization of glucose by brain tissue is made evident by its addition to brain slices, an increased consumption of oxygen then occurs. The oxygen consumption of brain tissue *in vitro* is also increased by the addition of fructose, mannose, galactose, and hexose diphosphate and monophosphate, and to a much less extent by other sugars. Next to glucose, fructose and mannose are most readily oxidized by brain slices. Lactate and pyruvate are utilized about as

readily as glucose. Glutamic acid and alcohol are utilized by brain tissue to a small extent.

In hypoxia and during convulsive seizures, large quantities of lactate are produced by brain tissues, but as already mentioned, only minimal amounts when the oxygen supply is adequate. The large production of lactate under anaerobic conditions and the minimal amounts produced when the oxygen supply is adequate raises the question whether the formation of lactic acid plays any rôle in the normal function of the brain. It appears, at any rate, that it does not play an indispensable part in the oxidation of glucose, for nicotine, iodoacetate or hydroxymalonate inhibits the oxidation of lactate but permits glucose to be oxidized. During convulsions, or when from any cause the oxygen supply is inadequate, the brain, like muscle, uses phosphocreatine. Similar changes are produced by injury to brain tissue. Though other fuel can be oxidized by the brain, glucose, lactic acid and pyruvate are the only ones, and especially glucose, which are utilized normally to any significant extent. In excised brain tissue "the utilization of glucose is faster than that observed in hepatic, cardiac, or renal slices" (Himwich). A respiratory quotient of unity supports the conclusion that the main fuel of the brain is carbohydrate, but more direct evidence for the oxidation of glucose can be cited. Himwich and Fazekas, for example, determined the glucose and oxygen contents of the arterial blood and of blood drawn from the superior longitudinal sinus of anesthetized dogs. They obtained a figure of 13 mg. per 100 cc. of blood for the glucose utilization. The quantity of oxygen which by calculation would be required to oxidize this quantity of glucose is 9.7 cc. The oxygen consumption actually observed was 9.3 cc. These figures are, of course, for the brain as a whole.

The gray matter shows a much higher oxygen consumption than does the white matter which is composed mainly of nerve fibers. Though the nerve fiber can contract a small oxygen debt, the gray matter cannot and, in consequence, is highly susceptible to oxygen deprivation. It has been estimated that the adult brain accounts for about one-quarter of the basal metabolism of the entire body at rest, or over 3 liters per hour, yet mental work increases the metabolism of the brain to a negligible extent (ch. 45). The infant brain which, as compared with the adult brain, forms a larger part of the body has a relatively greater consumption of oxygen, its oxygen consumption is about

half the total oxygen consumption of the body. Considering the high oxygen requirement of the gray matter, especially of the cerebral cortex, it is not surprising that it is so vulnerable to anoxia. Mental confusion, delirium and unconsciousness follow any considerable reduction in the oxygen supply to the brain. The delirium of pneumonia is due largely to anoxia rather than to toxemia, and the mental aberrations associated with high altitudes are well known. The metabolism of the cerebellar cortex, according to some investigators, is higher than that of the cerebral cortex. Himwich and Fazekas have made the interesting observation that in week-old puppies the medulla and midbrain have a higher oxygen consumption than has the cerebral cortex but the reverse relationship holds true in adult dogs. The increased oxygen consumption of the higher centers of grown animals is due presumably to their having acquired greater functional importance and assumed a position of dominance over the lower centers.

The very high oxygen consumption of gray matter as compared to white matter (nerve fibers) is not due to the presence of cell-bodies, but apparently to the great number of synapses (ch 64), for the posterior root ganglia, which are made up of unipolar cells without synaptic connections, have not a high vascularity or oxygen consumption, and are not more susceptible to ischemia than are nerve fibers.

The brain is also very sensitive to a reduction in its supply of glucose, and when the blood sugar level falls to a certain point, mental confusion, muscular incoordination, convulsions and loss of consciousness result. The hypoglycemic symptoms are quickly relieved by the administration of glucose, fructose or mannose but not by pyruvate or lactate. In the human subject insulin hypoglycemia greatly reduces the oxygen consumption of the brain, but this is quickly restored to normal by the injection of glucose (Himwich and associates). The reduction of the metabolism of the brain and its restoration to normal are closely associated in time with the hypoglycemic symptoms and their relief. Yet, strangely enough, the brain continues to utilize glucose in the absence of insulin, the respiratory quotient remains at unity and the arteriovenous oxygen difference is unaltered after pancreatectomy.

As might be expected from the relationship between the blood glucose level and the oxygen consumption, the effects of anoxia and of hypo-

glycemia are in certain respects closely similar, and are supplementary to one another. Hypoglycemic convulsions are induced more readily in the presence of anoxia, and the effects of the latter are more severe if the blood sugar is depressed.

There is no evidence that the lipids—*cholesterol*, *lecithin*, *cephalin* and *sphingomyelin*—which enter so largely into the composition of brain substance serve any special metabolic need, their function is probably to insulate neighboring nerve fibers from one another. Nor are fats or amino acids, with the exception of glutamic acid, utilized by the brain. This amino acid is capable of restoring consciousness to a subject of hypoglycemia, it is probably oxidized directly by the brain, but may also cause, through an adrenaline-like action, the liberation of glucose from the liver. Glutamic acid serves as a co-enzyme for the action of choline acetylase (ch 72). Acting upon the possibility that acetylcholine synthesis was depressed in petit mal epilepsy, glutamic acid has been employed in the treatment of this disease.

THEORIES OF THE ACTION OF ANESTHETICS AND NARCOTICS

There has been much speculation as to the manner in which certain drugs induce anesthesia or narcosis, but no theory is entirely satisfactory. There are three main theories, which may be termed the *lipid solubility*, *membrane permeability* and *metabolic theories*. The lipid solubility (Meyer-Overton) theory is concerned mainly with the manner in which the anesthetic enters the neuron, rather than with its mode of action upon the function of nervous tissue after it has penetrated the cell. This theory is founded upon a certain, but far from perfect, correlation between the action of anesthetic agents, and their solubility in lipid (which the cell membrane and other structures of the cell contain) but low solubility in water. The narcotic effects of the chemically inert gases, nitrogen, argon, krypton and xenon are attributed to their ready solubility in lipid.

It was proposed by Lillie that the permeability of the cell membrane, with the consequent prevention of depolarization, upon which excitability depends, was reduced by anesthetic drugs.

The metabolic theory is based largely on the work of Quastel and Wheatly, who have shown that many narcotic drugs, notably the barbiturates, paraldehyde and urethane, depress the oxygen consumption of brain tissue *in vitro*, and reduce

the oxidation of glucose. The metabolic activity of the brain *in vivo* is also inhibited in anesthesia. These inhibitory effects upon brain metabolism are dependent, presumably, upon the interference with the function of some oxidative enzyme system, possibly by blocking the interaction between flavoprotein and cytochrome *b*. An important objection to this theory is the failure to establish a satisfactory correlation between the concentrations of drugs which are effective in inhibiting oxidation *in vivo*, and the concentrations which cause anesthesia. Ether, for example, causes insignificant effects upon the oxygen consumption of brain slices in concentrations which induce general anesthesia.

It is possible, however, that the concentrations of anesthetic drugs in the brain as a whole may be much lower than in certain special and restricted regions.

One point seems to be generally agreed upon with respect to anesthetic and narcotic agents, namely, that they act through a physical rather than through a chemical mechanism. All the evidence points in this direction. No consistent chemical structure is associated with their anesthetic or narcotic property, their actions are reversible, they do not appear to enter into any chemical reaction, and they can be recovered almost quantitatively from the tissues unchanged.

CHAPTER 63

THE REFLEX ARC RECEPTOR ORGANS CUTANEOUS AND KINESTHETIC SENSATIONS

The involuntary muscular contraction which results from the stimulation of a sense organ is known as a reflex. The quick withdrawal of the hand from some agent which has inflicted pain or the contraction of the pupil when a light is thrown into the eye are familiar examples of reflex action. The activities of various glands are also largely reflex in nature as are many of the reactions of the vascular, respiratory and digestive systems. The exciting cause of the reflex may make an impression upon consciousness, as when a group of skeletal muscles contracts as a result of painful stimulus, or, as in the case of reflexes involving the secretion of glands or the activities of smooth muscle, e g , of the blood vessels, heart or digestive tract, the initiating stimulus as well as the reflex act itself may be entirely unperceived.

THE REFLEX ARC

The anatomical basis of reflex action is the reflex arc which in its simplest conceivable form consists of

(a) *An afferent limb* composed of the *receptor organ*, which upon excitation, gives rise to the impulse, and the *neuron* whose processes (central and peripheral) transmit the impulse to the central nervous system. In the case of the spinal reflex arc the cell bodies of the afferent neurons are situated in the posterior root ganglia.

(b) *An efferent limb* constituted of a motor or a secretory neuron which conducts impulses from the central nervous system to an *effector organ*—muscle or gland. In the case of motor spinal reflex arcs, the axons of the efferent neurons leave the cord by the anterior nerve roots and travel in the peripheral nerves, their cell bodies are situated in the anterior horns.

(c) *A center* situated in the gray matter of the central nervous system and consisting of the cell body of the efferent neuron and its junction (synapse) with the central process of the afferent neuron.

As a rule the afferent and efferent limbs do not connect directly in the center, one or more nerve cells being interposed between the two. These are spoken of as *connector*, *internuncial*, or *inter-*

calated neurons (fig 63.1). The stretch reflex (p 960) as shown by Lloyd is carried out through a reflex arc of only two neurons, but the great majority of the spinal reflex arcs in higher animals consist of several neurons, and in most reflexes each afferent neuron makes connection through collateral branches and internuncial neurons with a large number of motoneurons (fig 63.2).

Injury, leading to loss of function of any one part of the reflex arc, is sufficient to destroy the function of the whole.

THE RECEPTOR ORGANS

The afferent fibers end peripherally either as bare unmyelinated filaments or in accessory structures called *receptors* (fig 63.3). These are highly specialized to respond most effectively to one or other type of stimulus. When stimulated appropriately an impulse or a series of impulses is sent along the afferent fiber. Receptors are situated in the skin, muscles, tendons, etc., and in such special organs as those of sight, hearing, smell and taste. They are also contained in the walls of the respiratory and digestive tracts, mesentery, carotid sinus and other internal structures. Through receptors in these various situations messages (nerve impulses) are continually being transmitted over somatic and autonomic pathways to the central nervous system.

Those receptors which respond to stimuli arising in the outer world, e g , in the skin, eye, ear, etc., are called *exteroceptors*. Of these, ones which make perception at a distance possible, i e , those situated in the visual, auditory or olfactory sense organs, are sometimes referred to as *distance receptors* (telereceptors). Receptors lying in the mucous linings of the respiratory or digestive tracts and which, though not in immediate contact with the outer world, respond to stimuli ultimately derived therefrom are spoken of as *interoceptors*. *Proprioceptors* are those which respond to stimuli originating within the body itself, e g , in the skeletal muscles, (p 942) tendons, joints, heart, carotid sinus, gastro-intestinal wall, etc. Though each variety of receptor responds most readily to one particular type of

stimulus—adequate stimulus—many will respond in some degree to stimuli of other types. The retina, for instance, can be stimulated mechani-

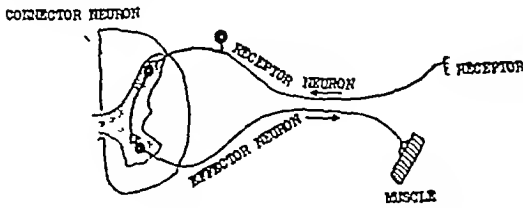


FIG 63.1 Diagram of a simple reflex arc.

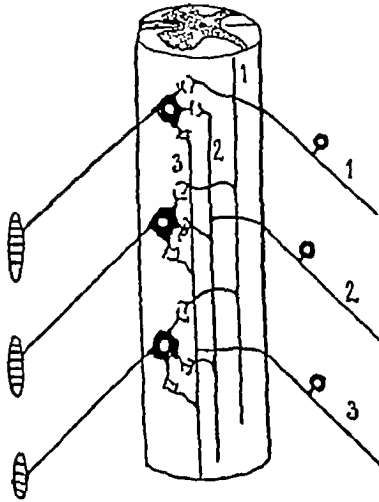


FIG 63.2 Showing how an afferent nerve fiber upon entering the cord makes connections with several motor neurons, and how each of the latter is in communication with several afferent fibers. Connector neurons are not shown.

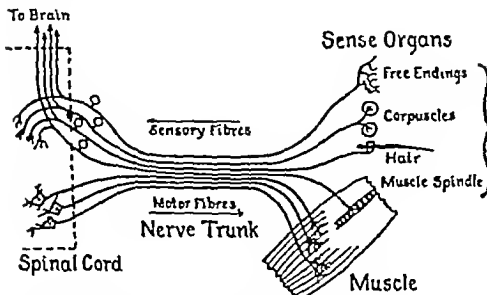


FIG 63.3 Diagram to show the nervous connections between the central nervous system and the periphery (muscle and cutaneous receptors) (from Adrian, *The Basis of Sensation*)

cally or electrically and the receptors of taste, though responding most effectively to chemical stimuli, may also be stimulated by an electric

shock applied to the tongue. A terminology based upon the type of stimulus which excites them adequately is sometimes used to designate different varieties of reception organ. Thus, *tango-* (touch), *chemo-* (taste, smell and the receptors of the carotid and aortic bodies) and *photo-* (sight) receptors are spoken of.

CUTANEOUS SENSATIONS

The sensations which may be aroused by stimulation of the skin are *touch*, *cold*, *warmth* and *pain*. Each of these sensations, except the last, is mediated by a receptor or sense organ possessing distinctive structural features.

The presence of discrete endings subserving the several cutaneous sensations enables small areas to be mapped out upon the skin which are specific for one or other sensation. The areas are called touch, cold, heat or pain "spots" respectively in accordance with the sensation which their stimulation arouses (fig 63.4).

If the *afferent nerve fiber* supplying one or other receptor organ is excited *directly* by the application of the type of stimulus, e.g., touch, heat, etc., for which the receptor itself is adapted to respond, the characteristic sensation is not, as a rule, aroused, a painful sensation usually results.¹ Moreover, reflexes can usually be elicited much more readily by stimulating the receptors than by applying the stimulus directly to the afferent nerve, and certain reflexes cannot be evoked at all by direct excitation of the nerve fiber. Pressure upon the pad of the hind foot of the "spinal dog", for example, causes a strong extension of the whole limb—the *extensor thrust*—whereas no form of stimulus applied directly to the afferent nerve itself will produce this reflex (Sherrington). The afferent fiber before terminating in the receptor organ or as a free nerve filament loses its myelin sheath and neurilemma and appears as a naked axis cylinder.

LIGHT TOUCH. Tactile sensation—the sensation aroused by light contact—is subserved by three types of receptor, *Meissner's corpuscles*, *Merkel's disks* and a basket-like arrangement of nerve

¹ Heinbecker, Bishop and O'Leary have reported that tactile as well as painful sensations may be aroused by the electrical stimulation of the nerve trunk. The experiments were carried out upon the exposed nerves of human subjects. Stimuli of low intensity applied to the nerve aroused the sensation of touch, those of greater intensity a sensation of pain. No other type of sensation was experienced by the subjects.

fibers, surrounding the base of hair follicle (fig 63.5) Meissner's corpuscles are situated in the papillae of the skin, just beneath the epidermis. They are unevenly distributed, being sparsely scattered over such a region as the volar aspect of the forearm but numerous in the skin of the hand, foot, nipple and lips, and in the mucous membrane of the tip of the tongue. They are well organized structures, consisting of irregularly

short, vertical, nerve filaments which end in small bulbous expansions. They are stimulated by any slight movements of the hairs (See fig 63.5F)

Inequality of pressure with consequent deformation of the skin surface is the essential factor in the stimulation of touch receptors. If the pressure is distributed equally, and, therefore, no deformation of the skin occurs, a tactile sensation is not experienced, when, for example, a finger is dipped into mercury a sensation is aroused only in the narrow band of skin where it is deformed as a

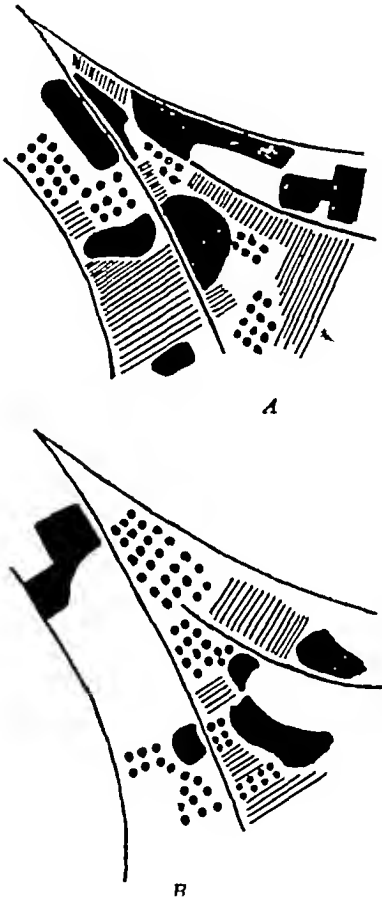


FIG 63.4 Showing cold spots (A) and hot spots (B) within an area on the palm of the hand. The sensation in each case was most intense in the black areas, less intense in the lined and mildest in the dotted areas. In the blank portions no definite sensation was aroused (after Goldscheider)

coiled nerve endings with capsules of connective tissue. *Merkel's disks* consist of groups of three or more cup-shaped disks with a reticulated appearance. The nerve fiber upon approaching a group of such structures breaks up into branches, one going to each disk. *Merkel's disks* are found in the skin of the snouts of pigs and other mammals and in the finger-tips, lips and mouth of man. The basket-like arrangement, surrounding the base of a hair follicle consists of a number of

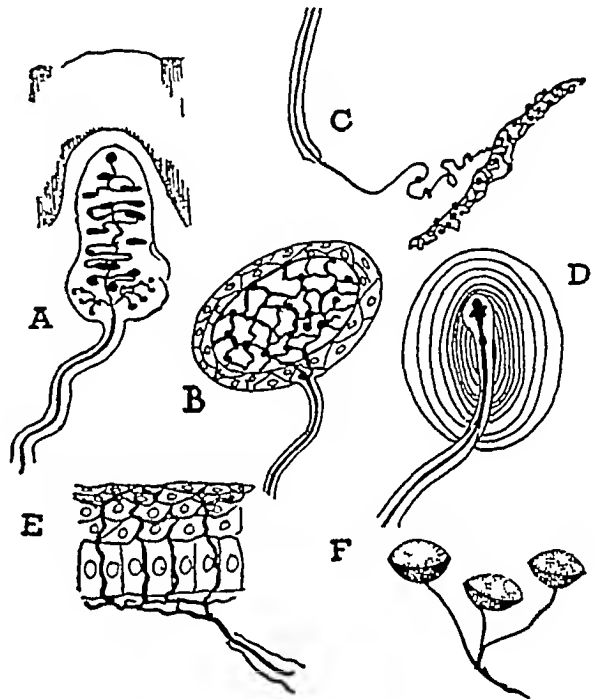


FIG 63.5 Cutaneous receptors. A, Meissner's corpuscle (touch), B, Krause's end bulb (cold), C, Ruffini's end organ (warmth), D, Pacinian corpuscle (deep pressure), E, bare nerve endings in cornea (pain), F, Merkel's disks (touch) (in part from Bainbridge and Menzie, *Essentials of Physiology*, Longmans, Green and Co)

result of unequal pressures at the interface between the mercury and air

The sensation of light touch is tested by bringing a wisp of absorbent cotton in contact with the skin, or by the use of von Frey's esthesiometers. These consist of a series of hairs of graded thickness attached at right angles to wooden holders. The pressure in grams required to bend each hair is known. In order, therefore, to express the sensitivity of the skin to touch in terms of pressure the hair is found by trial which, when pressed vertically upon the skin until bending occurs,

causes the sensation. The sensitivity of the skin to touch varies widely in different regions. The minimal pressures required are given in table 88. When hairy parts, such as the back of the hand, are lightly brushed with a tuft of cotton wool the hairs serving as levers deform the skin, and thus cause stimulation of touch receptors situated in the neighborhood of the hair follicles. Shaving the hairs over such parts greatly reduces the sensitivity to touch. Other regions quite devoid of hairs such as the finger tips and lips, on the other hand, possess the highest degree of tactile sensibility.

TACTILE LOCALIZATION. When a tactile stimulus is applied to a point upon the skin the normal

TABLE 88

The minimal pressures required for the detection of the sensation of touch from various cutaneous regions

(After Meyers)

REGION	GRAMS PER SQUARE MILLIMETER
Nose	2
Lips	2.5
Tip of finger	3
Back of finger	5
Upper arm, inner surface of thigh	7
Back of hand	12
Calf, shoulder	16
Abdomen	26
Front of leg, sole of foot	28
Back of forearm	33
Loam	48

subject is capable of recognizing the location of the stimulus with a high degree of accuracy. This faculty is termed *topognosis*. A tactile sensation has, thus, associated with it a localizing quality in addition to its specific characteristic, which has been called "local sign." Though there is a neural basis for point to point projection of skin areas to the cerebral cortex, the actual faculty of locating the point touched appears to be acquired by experience and not inborn. Localization is much more precise over some regions, such as lips and tips of the fingers, than over others, such as the forearm or thigh. In some nervous diseases this localization is grossly impaired. In certain lesions of the cerebral cortex the subject usually, according to Horsley, when asked to locate the stimulus indicates a

point some distance from it on the *proximal* side. In some diseases, notably hysteria, the subject feels the stimulus at a corresponding point on the opposite side of the body, or in a limb on the side opposite to that to which the stimulus was applied, this phenomenon is called *allochæria*.

The localization of a cold, hot or painful sensation is very inaccurate unless the stimulating agent actually touches the skin. Thus, heat radiated from a small object about 1 millimeter from the cutaneous surface gives rise to a diffuse sensation. The more accurate localization when contact is made with the skin is evident though the subject experiences no sensation of touch. It is likely, nevertheless that the tactile receptors are excited, that the failure to appreciate touch is due to masking by the stronger stimulus, and that impulses arising in the touch endings are responsible in some manner for the more accurate local-

TABLE 89

Different cutaneous areas compared with regard to the minimal distance which must separate two stimulated points in order to arouse a double sensation

(After Meyers)

REGION	MINIMAL DISTANCE MM.
Volar surface of finger tip	2.3
Dorsal surface of third phalanx	6.8
Palm of hand	11.3
Sole of foot	16.0
Back of hand	31.6
Back of neck	54.0
Middle of back, upper arm and thigh	67.1

ization of the other types of stimulus when these are applied directly to the skin.

TACTILE DISCRIMINATION (TWO-POINT SENSIBILITY, COMPASS TEST). If two stimuli are applied simultaneously, two distinct sensations are felt, provided the distance between the two stimulated points is sufficiently great. Thus, when the points of a pair of compasses are blunted or covered with cotton wool and applied to the finger tip, the subject recognizes the duality of the stimulus if the points are more than about 2.3 mm. apart. When they are separated by a shorter distance a single sensation is experienced. The minimal distance at which the recognition of two stimuli is possible varies in different regions, as shown in table 89.

In the case of the limbs the power of discrimination diminishes progressively from the distal to

the more proximal segments, and hairless regions in general have a higher discriminating ability than those covered with hair

A correspondence is also exhibited between the mobility of a part and its discriminating ability. For example, the minimal distance necessary for two stimuli to give rise to a double sensation is less for the fingers and hand than for the arm, shoulder and back, and diminishes progressively over the skin of the face from the region of the ear to the lips

It should be pointed out that the values given in the foregoing table do not represent the distances separating individual touch receptors. For example, when single stimuli are applied *successively* to different points of the skin of the finger tip the spots from which tactile sensation can be evoked are found to be about 0.1 mm apart. That is, when compass points are applied to the finger tip and are recognized as two stimuli several (15 to 20) touch spots are included within the intervening space

TICKLING AND ITCHING There has been much discussion concerning the origins of these sensations. The *tickling* sensation caused by light stimulation of the skin, as by a straw, appears to be due to the summed effects of stimulating both touch and pain endings. Section of the lateral spinothalamic tract (which conveys pain impulses) results in the loss of the appreciation of pain but retention of the sense of touch, a tickling sensation cannot be aroused over the analgesic skin. On the other hand, the sense of touch is lost while that of pain is retained, when the skin is rendered moderately ischemic, again, the tickling sensation cannot be elicited. The *itchiness* which is experienced in the region adjacent to a slight injury, or during the healing of a more severe injury when the skin is rubbed, has the same nervous origin, being dependent upon impulses travelling by both tactile and pain fibers, but the *spontaneous* itching (i.e., in the absence of any external stimulus) which is felt under the same circumstances, appears to be due to the mild stimulation of pain endings alone, more intense stimulation of the same character causes pain. This type of itchiness is affected relatively little by ischemia. Both types of itching are alike in that they are due to a chemical substance acting upon nerve endings and liberated by the damaged cells of the skin. This substance appears to be the same as that which causes the triple response. An extract of skin showing this

response when injected into normal skin induces both types of itching. Histamine introduced into the skin has a similar effect. Rothman believes that itching is identical in quality with, and varying only in intensity from protopathic pain and is mediated by the C group of nerve fibers

Cold is mediated by the end organs of Krause (fig. 63.5 B) and *warmth* probably by the end organs of Ruffini (fig. 63.5 C).² Mechanical or electrical stimulation, as well as the application of heat itself, will stimulate the latter end organs and give rise to the sensation of warmth. The receptors for warmth (Ruffini's) are situated in close proximity to the deep plexus of blood vessels, when these latter dilate the endings are stimulated by the warmer blood coming from deeper regions. The cold receptors are also stimulated at temperatures above 112°F giving rise to a sensation, not of heat but of cold. Cold, of course, is a negative quality—the sensation being due to the withdrawal of heat. It would be more correct, therefore, to say that the cold receptors are stimulated at extremes of temperature, above 112°F or below that at which the warm receptors respond. Cold and touch spots are much less numerous than touch and pain spots

PRESSURE upon the skin considerably greater in degree than that which elicits the sensation of touch stimulates the more deeply lying receptors known as the *Pacinian corpuscles* (fig. 63.5 D). The sensation of pressure is not, however, a true cutaneous sensation, the Pacinian corpuscles are situated in the subcutaneous tissues or inner

² Waterston and others have questioned the existence of definite heat and cold spots since it has been found that the points from which these sensations can be aroused change from time to time in number as well as in pattern. He states that when the skin is hyperemic the entire surface responds to warmth (punctate distribution being abolished) and he believes that the entire skin surface is potentially sensitive to this sensation, the apparent punctate distribution being due simply to fluctuations in the activity of different areas. The variability in sensation, it has been suggested may possibly be related to a corresponding fluctuation in the capillary circulation. The observations of Bazett and associates indicate, however, that the diffuse nature of the temperature sensations in hyperemic skin can be best explained upon the basis of a more ready conduction of heat through the blood stream to neighboring end organs, thus, when the cutaneous blood flow is profuse, heat applied to a nonsensitive point in the skin would be conducted rapidly to adjacent warm spots. The observations of Bazett and his colleagues upon the sensitivity of the prepuce support the conception of heat and cold spots and the existence of discrete receptor organs for these sensations as well as for touch (Arch. Neur. Psych. 1932, 27, 489)

layers of the dermis as well as in tendons, periosteum and other deep seated structures, their nerve fibers run chiefly, not in the cutaneous nerves, but in the sensory fibers of mixed nerves supplying muscles, tendons, joints and blood vessels. Since the sense of deep pressure is preserved after sectioning the cutaneous nerves, the loss of light touch, a purely cutaneous sense, may, therefore, be undetected in a lesion affecting this sensation, unless the contact is made very lightly, as by a wisp of cotton wool, with the skin, otherwise receptors in the subcutaneous tissues may be stimulated.

PAIN is subserved by naked nerve endings, there being no organized end organ for this sensation. The pain nerves of the skin are described by Woolf as consisting mostly of non-medullated fibers which terminate in the superficial layers of the dermis in delicate loops lying parallel to the skin surface, or as long naked neurofibrillae. Only occasionally do fibrils penetrate the epidermis. Bare nerve endings mediating pain are also present in the cornea (fig 63.5 E) and in serous surfaces (peritoneum, pleura, etc.), touch, cold and warm endings are absent and the corresponding sensations cannot be aroused from these locations. Although it has been the general belief that the cornea contains only pain endings and that any stimulus, if intense enough to evoke a response at all, causes pain. The sensation of cold can, however, be elicited from the cornea, and according to some, the sensation of touch without pain can be aroused by a very mild stimulus such as a jet of isotonic saline. Certain structures, e.g., the tooth pulp, the middle meningeal artery, the arteries at the base of the brain and some of the vessels of the scalp contain no sensory fibers except those which give rise to pain. Other structures, on the contrary, e.g., the substance of the brain and the mucosa of the cheek opposite the second lower molar tooth, are insensitive to pain.

Most of the pain impulses arising in the skin travel to the central nervous system in the somatic nerves, but some of those from deeper structures join autonomic nerves. All enter the central nervous system by the lateral divisions of the posterior spinal nerve roots or the cranial nerves.

The pain endings do not respond selectively to one variety of stimulus but to any type whether mechanical, chemical or thermal, provided it is sufficiently intense. The pain stimulus, whatever it may be, has one property in common, namely,

that it causes or threatens injury. The sensation of pain therefore serves a protective purpose, giving warning of the injurious nature of a stimulus rather than information as to any more specific quality. Stimuli which arouse painful sensations also provoke reflex actions which have the following features: (a) They comprise movements for *protection* or *defense*, or for the withdrawal of the part from the noxious agent. (b) They are *prepotent*, other less urgent reflexes being for the time inhibited. (c) They are *imperative*. Such reflexes are called *nociceptive*.

Whereas, the touch spots are more numerous toward the peripheral parts of the body, the pain spots are more profuse near the roots of the limbs. Thus, in the axillae, supraclavicular fossae and inguinal regions, their number is around 200 per sq cm, but there are no more than from 40 to 70 per sq cm over the palms and soles, the tip of the nose and the ear.

It used to be thought that the production of pain was not a function of any one type of nerve ending, but that any skin receptor if stimulated with sufficient intensity would give rise to the sensation. This conception has been refuted by modern work which indicates that the various types of cutaneous receptors are specific in function (see p 934). Interruption of the pain tracts (lateral spinothalamic) of the cord, for example, abolishes pain, but thermal and tactile sensations are retained. Also, in the experiments of Adrian, Cattel and Hoagland (p 944) pain was not aroused by maximal stimulation of touch receptors. Finally, certain analgesic drugs reduce sensitivity to pain, while actually lowering the threshold for touch and the discrimination of two points.

PERCEPTION OF AND REACTION TO PAIN

Pain thresholds. A distinction should be drawn between the *perception of pain* and the *reaction* which results, e.g., contraction of facial muscles, vocalization, narrowing of eyelids, changes in pulse rate or rise in blood pressure, sweating, and vasomotor responses. In man the *perception threshold*, under similar conditions, remains remarkably constant between normal persons, and for the same person from day to day, or from hour to hour of the day. The threshold is altered, however, by such extraneous influences as a loud noise, gripping some object or clenching the jaws, which may raise it by 40 per cent. The threshold may be definitely raised by a placebo, and some distraction, such as

pain or discomfort, in one situation may reduce the perception of pain in another. The twitch applied to a horse's nose during a minor operation is a familiar example. Even other sensations, e.g., touch, pressure (rubbing), warmth, etc., may raise the pain threshold. The analgesic action of counter-irritation is, in some instances, dependent upon such an effect. Hypnosis also raises very considerably the perception threshold, as do alcohol and, of course, other analgesic drugs.

The *threshold of reaction* to pain varies widely between different subjects. In those of a stoical, phlegmatic temperament, in prize-fighters, Negroes and North American Indians the threshold is high whereas in "high-strung", neurotic persons it is low. Alcohol and analgesic drugs raise the reaction threshold to a greater degree than they do the perception threshold (Wolff)³

Discrimination of pain intensity When the strength of a pain stimulus is increased gradually from the threshold to double the threshold value, it is found that the intensity of the sensation increases in a series of 22 "steps" or just perceptible changes in intensity.⁴ This is a very fine discrimination as compared with other sensations.

The nature of cutaneous pain Pain in the skin, although usually described as of several different kinds, e.g., burning, pricking, cutting, etc., is always, according to Lewis, of the same quality, the apparent differences being due solely to variations in the duration of the sensation. A blind-fold subject, for example, is unable to distinguish between the pain caused by a pin prick, a hot point, a punctate electrical stimulus or the plucking of a hair, provided no associated non-painful sensation gives the patient a clue as to the nature of the stimulating agent. These pains would all be described as sharp, "bright" or pricking. A burning pain is experienced when the sensation is more prolonged, whether caused by heat or ultra-violet light or by a chemical or mechanical

irritant. There is no spatial summation of pain stimuli, i.e., the threshold for pain is not influenced by the extent of the area stimulated. Nor, in contrast to stimulation by touch or warmth, adaptation to a painful stimulus does not occur, pain continues to be felt as long as the stimulus is applied. The biological significance of this is obvious.

It is a common experience that a single painful stimulus applied to the skin, if intense enough, may give rise to two sensations separated by a short interval. The first pain is short and sharp, the second, more prolonged and severe. The observations of Lewis and Pochin upon this phenomenon indicate that pain impulses are conveyed from the skin by two sets of nerve fibers, one of which is rapidly conducting, the other with a much slower conduction rate. They found that ischemia of the skin of the arm, induced by arresting the circulation, abolished the first response to the needle prick, the second response being unaltered for a time but later became reduced. Cocaine, on the other hand, abolished the second response before the first. It was shown moreover, that the time interval between the two responses was prolonged as the length of the nerve between the point of stimulation and the central nervous system was increased, a fact which seems to demonstrate decisively the existence of a fast and slow pathway for the transmission of cutaneous pain. The time interval between the two responses is 1.9 sec. when the stimulus is applied to the toe, 1.3 sec. to the knee and only 0.9 sec. to the upper limit of the thigh.

The first painful sensation of the dual response appears to travel by the fibers of rapid conduction belonging to Erlanger's and Gasser's A group (p. 925) with a conduction rate up to 120 meters per second, they are more susceptible to asphyxia, the second pain response is transmitted by the slowly conducting C group of fibers (conduction rate of less than 2 meters per second), they are more readily affected by cocaine.

The rapidly transmitted pain appears to be the sensory basis for protective reflexes. Pochin found that the "fast pain" is abolished in tabes but the "slow pain" is retained. In this disease the response to pin prick is delayed and the withdrawal reflex also abolished.

A third type of pain response follows some little time after certain forms of injury, e.g., scorching, scalding, sun-burn or the application of an irritating agent, and persists for a variable period. Even a slight burn may cause a dull pain which con-

³ In determining the pain threshold in man, the Hardy-Wolff-Goodell method may be used. In this method radiant heat from a 1000-watt projection lamp, focussed for 3 seconds on a blackened area on the subject's forehead is used as a stimulus. The current passing through the lamp filament is gradually increased until the sensation of warmth gives place to a sharp "jab" of pain at the end of the 3 second period. The intensity of the applied stimulus is expressed in millicalories/sec/cm². The perception threshold is indicated verbally by the subject, and the reactivity threshold by a change in skin resistance as determined by means of a Wheatstone bridge and galvanometer.

⁴ Two steps are taken as a unit of pain and designated one *dol*.

tinues for many minutes. This pain evidently is caused by the release of a chemical substance from the damaged tissues and not by direct action of the stimulating agent upon the nerve terminals. Arresting the blood supply to the part intensifies and prolongs the pain. The nature of the chemical excitant is unknown, but it is not H substance (p. 316), potassium, nor acetylcholine, nor is it due to altered pH.

Dissociation of cutaneous sensations. In disease, the several modalities of sensation, touch, cold, warmth, and pain may be lost separately, or they may be temporarily dissociated by artificial means as by asphyxia, cocaine or the application of cold. Thus, the cutaneous sensations, touch, cold, warmth and pain are lost in this order when the skin is made ischemic. If the skin is cocaineized, the appreciation of cold is lost first, then follow in order the senses of warmth, pain and touch. Cooling the skin causes first, failure in the response of the cold receptors, then in succession the sensations of touch, pain and warmth are lost. Also, after section of a cutaneous nerve, the area of pain loss is smaller than that of touch, there is thus a boundary zone from which a response to a painful stimulus, such as that of a pin prick, can be obtained but which is insensitive to light touch.

Deep sensibility. Pain in the deep structures, muscles, bones and joints, etc., the sense of pressure in the subcutaneous tissues, and sense of position and movement (e.g., kinesthetic sense) of joints come in to the category of deep sensibility. Deep sensibility is mediated by afferent fibers in the mixed nerves. Pain arising in deep structures differs in certain respects from superficial pain. It is poorly localized (ch. 44), and is often of a dull, aching or "sickening" character as compared with superficial pain which tends more usually to have a "bright", sharp or burning quality. Deep pain is more often accompanied by nausea and vomiting, slowing of the pulse, and a fall in blood pressure, whereas cutaneous pain is more commonly associated with a quickening of the pulse and a rise in blood pressure.

Muscles, tendons and fasciae are especially susceptible to painful stimulation by chemical agents. The injection of a few drops of a 6 per cent saline solution into one of these structures causes pain. Muscle is relatively insensitive to pricking or cutting but pain is aroused by pressure, e.g., pinching or squeezing, or by exercising under ischemic conditions (p. 300). Tension acts also

as a pain stimulus for muscle, tendon or fascia. Pain occurring in ischemic muscle during activity is due to a chemical irritant produced by the active tissues and which accumulates and stimulates the pain endings when the circulation to the part is arrested or considerably reduced. This substance is referred to as factor P by Lewis and his associates. The soreness of healthy muscles which comes on some hours after exercise is of the same nature.

Periosteum and cancellous bone are very sensitive to the various types of mechanical stimulation, but compact bone is insensitive to drilling or sawing. The arteries give rise to painful sensations when pricked, but the walls of the veins, except the larger intracranial veins, are usually insensitive.

Visceral pain and referred pain are dealt with in ch. 44, *central pain* in chaps. 67 and 68, and *headache* in chap. 68.

Hyperalgesia and the nocifensor nerves. In many persons an area of tenderness develops around even a small cutaneous injury and spreads for a considerable distance in all directions. The soreness starts within a few seconds, increases to a maximum in from 15 to 30 minutes and lasts for hours or, with a more severe injury, for days. The threshold for pain as tested by a needle prick is lowered only slightly over the area, but the pain when aroused is diffuse, and unusually intense and prolonged. This phenomenon has been studied by Lewis and his associates. Injury was produced by prolonged faradic stimulation of the skin, by crushing a small cutaneous fold with forceps or by direct stimulation of a cutaneous nerve trunk or one of its branches. A similar area of cutaneous hyperalgesia may result from an injury or an inflammatory process involving deep-lying tissue or mucous membrane. Thus, stimulation of a dental nerve or of the mucous membrane of the maxillary antrum is followed by tenderness of the overlying skin of the cheek. Lewis' experiments point to a specific system of nerves in the skin as being responsible for the hyperalgesia. They are believed to be quite distinct from the pain fibers which are stimulated directly by the injurious agent and through which pain sensations are ordinarily registered. This system, which is referred to as the *nocifensor nerves* is pictured as a finely branching and rich nerve plexus within the skin, it leads into more deeply lying fibers which, in turn, run into subcutaneous nerve trunks. The action of this system of nerves serves to ward off further injury and to put the injured part at rest, hence their name. They are not sympathetic

filaments, for the phenomenon is observed in skin completely deprived of its sympathetic innervation. Lewis bases his conclusion that the nerves which respond in the ordinary way to pain stimuli are not responsible, upon the diffuse and poorly localized character of the hyperalgesia, and upon the effects of ischemia and of cocaine. The nerve fibers responsible for the hyperalgesia are paralyzed by ischemia but are little affected by cocaine. Pain sensations, on the contrary, are affected relatively little by ischemia but are readily abolished by cocaine.

Epicritic, protopathic and deep sensibilities

Head and his associates, from the study of a large number of peripheral nerve injuries, grouped the several superficial sensations into two classes: (1) *epicritic*, and (2) *protopathic*. The sensations grouped under *epicritic sensibility* included light touch over hairless parts, the power of localizing the point touched, the detection of two individual sensations when two points are touched simultaneously, and the appreciation of finer grades of temperature, cool and warm, i.e., temperatures ranging between about 40° and 25°C. *Protopathic sensibility* is a more primitive type of sensation and more widely distributed. It includes pain and the temperature sensations aroused by extremes of heat and cold—above 40° to 50°C and below 20° to 25°C. In other words, protopathic sensibility possesses a high threshold, but, though the stimulus must be strong in order to arouse a sensation, this once aroused is intense, diffuse, poorly localized and peculiarly unpleasant in quality. These latter qualities of protopathic sensibility are particularly prominent in the absence of epicritic sensibility, which therefore has been considered to exert a restraining influence upon the former. The glans penis possesses only protopathic and deep sensibilities.

For the purpose of studying these two types of sensation Head underwent an operation in which his radial nerve and the cutaneous branches of the musculo-cutaneous nerve were severed at the elbow. After the operation there was *complete* loss of cutaneous sensation over the radial half of the back of the hand. Surrounding this was a narrow zone in which epicritic sensibility alone was lost, protopathic was retained. This is due to overlapping of the protopathic innervation of the adjacent unaffected skin area. Over a small triangular area of skin at the wrist the relationship between these two sensations was reversed, protopathic sensibility was lost but epicritic retained. Stopford has since shown

that this latter type of dissociation of the two groups of sensations is also characteristic of section of a posterior root—the area of protopathic loss being much greater than epicritic. In Head's experiment evidence of returning sensation (due to regeneration of the nerves) was noticed on the 43rd day after operation when the zone of protopathic sensation was found to have become broader and to have encroached upon the totally insensitive area. The original area insensitive to light touch remained unreduced. Within about six months *protopathic* sensibility had returned over nearly the entire area, yet the extent of the *epicritic* loss remained unaltered. The first signs of the return of epicritic sensation was not observed until a year after the operation and was not complete after the lapse of two years. Deep sensibility was not affected by the nerve section referred to above, it is not mediated by fibers running in superficial nerves. Moderate pressure could be everywhere appreciated. This is an important point to remember in the investigation of nerve lesions since it shows that crude methods for studying the sensations of touch may fail to reveal any loss. The superficial touch receptors are stimulated by light contact with a wisp of cotton or by the pressure of a fine hair, whereas a pressure not very much greater, as with the point of a pencil, will stimulate the deeper lying pressure receptors (Pacinian corpuscles) whose nerve fibers may be intact while those of true touch are destroyed. As a result of these experiments in which dissociation of the three types of sensation were clearly demonstrated Head and his associates concluded that each sensation was mediated by a separate and structurally distinct group of fibers. Unquestionably, a deep sensibility is conveyed by fibers separate from those responsible for cutaneous sensations, but that epicritic and protopathic sensations has each its specific fiber group has been seriously questioned. The work of Erlanger and Gasser and of Ranson appeared to lend support to Head's interpretation. The C fibers of the former authors were suggested as the possible mediators of protopathic sensation and the A group, of epicritic. Ranson had previously described unmyelinated fibers in the cutaneous nerves which, since they were believed to mediate pain might offer an anatomical basis for the primitive protopathic sensations. The existence of protopathic sensation only in the glans penis and the dissociation of the two sensations, which as shown by Stopford result from section of the posterior roots, also supported Head's conception of two distinct sets of fibers. Experiments by other neurologists, on the other hand, have failed to substantiate this contention (see Trotter and Davies). The existence of two anatomically distinct sets of nerve fibers which could serve as a basis for the protopathic and epicritic types of sensation have never been demonstrated. Some would explain the two classes of cutaneous sensation following nerve section upon the basis of different rates of regeneration of the receptor organs, the ter-

minals of the pain fibers being thought to regenerate first. Fine discrimination, according to Heinbecker, is dependent upon the number of active end organs within a given area. The zone of protopathic sensibility surrounding an anesthetic area caused by nerve section is ascribed simply to the presence, in reduced numbers, of intact afferent endings which have overlapped the field of distribution of the sectioned fibers.

The recognition of these two types of sensation, whatever may be their anatomical basis, if indeed such exists, is, however, of distinct value from a clinical standpoint.

Vibration sense, pallesthesia This is the ability to perceive stimuli of a vibratory nature applied to the body surface, as by means of a low-pitched tuning fork, or other vibrating instrument. There is some doubt as to the character and situation of the receptors upon which this sense depends. According to most investigators they are present in bone, periosteum and tendons, and are therefore part of the system of deep sensibility, others contend that they are confined to the skin. Pollock has shown that the vibration sense is retained though the superficial sensations have been lost, which indicates that it is conveyed by fibers other than those mediating the latter sensations, namely by the deep system. Receptors for vibration sense are probably also situated in the deeper layers of the skin, and in the subcutaneous tissues. Though the sense is usually elicited by placing a vibrating tuning fork upon a bone, e.g. the shin or some other superficial bone, it can be aroused by applying the instrument to the soft tissues. The actual structures which respond to vibration are not known, but they are most likely the Pacinian corpuscles and the proprioceptors of tendon and muscle. The basis of this sense is probably the stimulation of receptors by very rapid changes in pressure, and has been compared to the rapidly repeated light flashes which arouse the visual sensation known as flicker.

Testing the vibration sense, whatever its anatomical or physiological basis, is an important diagnostic aid in neurological conditions. It is impaired or lost (and often before the sensation of touch or kinesthetic sense is affected) in lesions of the peripheral nerves, or of the posterior columns of the cord, but is not affected if both tactile and kinesthetic senses are intact—a fact that suggests that it is mediated by the same receptors which subserve both of these sensations.

Stereognosis (G. *sterio*, solid + *gnosis*, knowledge), *stereognostic perception*. This is the ability

to recognize, with the eyes closed, the size, weight and shape of objects placed in the hand or upon some other part. The absence of this sense is called *astereognosis*. When placed in the hand of a normal person while his eyes are closed, a coin, button, chain, ball or cube is easily identified from its size, weight, shape, texture (fur, wool, mesh) etc. Recognition may be made, though less readily, even if the object is laid upon the foot or pressed against the toes. Stereognosis is not a separate sense, but depends upon the senses of touch, spatial discrimination, cutaneous localization, and the kinesthetic sense. It is impaired or lost in lesions of the peripheral nerves or posterior columns of the cord, or of the somesthetic area of the cerebral cortex.

THE PROPRIOCEPTORS OF MUSCLES, TENDONS AND JOINTS

The receptors situated in skeletal muscles and in the tendons and joints furnish information to the central nervous system concerning the movements and positions of the limbs and other parts. Afferent fibers carrying this information make up from $\frac{1}{3}$ to $\frac{1}{2}$ of the fibers in a so called motor nerve. As a result of the messages received by the nervous centers the contractions of individual muscles and groups of muscles are coordinated to produce smooth, finely adjusted and effective movements which would be impossible in the absence of such guidance from the periphery. For this reason the term *kinesthetic* is applied to this group of receptors. A proportion of these afferent impulses arouse no sensation, their information being delivered to centers lying beneath consciousness. To others are due the sensations grouped in the section under deep sensibility. The receptors in the situations mentioned respond to mechanical stimulation, e.g., pressure or stretch. These types of stimulus are furnished by the strains and stresses set up in the muscles, tendons and joints during muscular contraction.

The sensory endings in the various situations mentioned above are of four main types: (1) *muscle spindles*, (2) *Golgi corpuscles*, (3) *Pacinian corpuscles*, and (4) *free nerve endings*.

(1) *The muscle spindle* is a fusiform body from 0.75 to 4 mm. long and from 0.1 to 0.2 mm. broad lying parallel to and between the muscle fibers (fig. 636). It is constituted of a bundle of from 3 to 10 muscle fibers (*intrafusal fibers*) enclosed in a fibrous capsule. The latter is separated from

the intrafusal fibers by a lymph space bridged across by delicate septa. The intrafusal fibers differ from the ordinary fibers of the muscle in being smaller and more circular on cross section, and in having a greater number of nuclei and coarser striations. The nerve supply of the spindle is double—afferent and efferent. The former enters the spindle about its center, the latter more toward one or other end. The efferent fibers are distributed to end organs in the intrafusal fiber which resemble but are not identical in structure with the motor end plates of ordinary muscle fibers. Upon entering the spindle and losing its myelin sheath and neurilemma, an afferent fiber may end in one or other of two ways: (a) Some become flat and ribbon-like and wind themselves in rings or spirals (*annulo-spiral endings*) about the intrafusal fibers. (b) Others ramify upon the

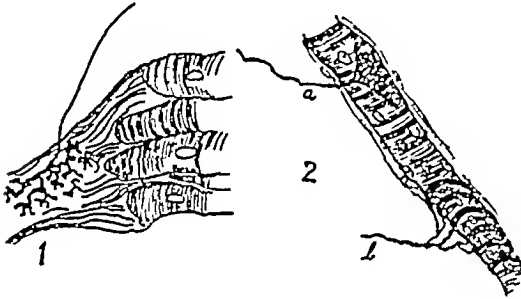


FIG 63 6 1 Golgi ending in tendon, 2 intrafusal fiber showing, a, muscle flower spray and b, annulo-spiral endings

latter's surface in a manner suggesting a spray of flowers ("*flower spray*" ending of Ruffini).

(2) *The Golgi corpuscles* (fig 63 6) are situated in tendons and consist of a bundle of tendinous fibers surrounded by a lymph space and enclosed within a fibrous capsule. Afferent nerve fibers enter the organ near its center and ramify upon its constituent fibers. Tension is the adequate stimulus for these receptors.

(3) *The Pacinian corpuscles* are oval bodies composed of concentric laminae, like the "skins" of a sectioned onion. The afferent fiber penetrates to the center of the corpuscle. Pressure with elongation of the organ and consequent stretching of the nerve ending is the adequate stimulus. These receptor organs are found in tendons, joints, periosteum, especially beneath tendinous insertions, in fasciae covering muscles and in subcutaneous tissues (fig 63 5). They are also found in the mesentery. The structure of these receptors appears to be such that a mechanical stimulus

such as stretching, pressure, etc., induced by muscular action will be applied most effectively to the naked axis cylinder within its center.

(4) *Free nerve endings* lie between the muscle fibers, in tendons and in the fasciae and joints. They mediate deep pain.

THE RESPONSES OF CERTAIN CUTANEOUS AND DEEP RECEPTORS

When non-polarizable electrodes are placed upon a mixed nerve of an intact animal at rest or upon a posterior root, a continuous stream of action currents can be recorded. These impulses of the "resting" nerve have a relatively low frequency—from 3 to 5 per second. They arise mainly from proprioceptors of muscle and tendons. The mere passive movement of the limb supplied by fibers entering the posterior root under investigation causes a prompt and pronounced rise in the impulse frequency. Light pressure upon the skin or even brushing the hairs also causes a rise in frequency as the result of stimulation of skin receptors. Adrian and his associates have recorded the afferent impulses from the frog's sciatic set up by stretching the gastrocnemius, from a nerve fiber supplying a single touch receptor of the frog's skin, and from the plantar digital nerve of the cat during the stimulation of the receptors of pain, touch or pressure in the skin of the toe pad.

When the receptors in the skin of the toe pad of the cat were stimulated by light contact or pressure the record taken from the plantar digital nerve showed a rhythmical discharge of impulses having a frequency of 150 per second. The frequency rose with an increase in the degree of pressure. Adrian and Umrath were able to locate a single Pacinian corpuscle and stimulate it by pressure with a hair or a fine glass rod. The impulses were recorded by means of an amplifier system in conjunction with a Matthews oscillograph. Pressure upon the end organ caused a discharge of impulses at a rate as high as 100 per second. A group of corpuscles lie beneath the flexor muscle of the toe, movement of the phalangeal joint crossed by the muscle stimulated the receptors and caused a discharge of impulses. Temperature changes failed to excite. The response of a single muscle spindle to stretch was also investigated by Adrian and Zotterman. They employed the sterno-cutaneous muscle of the frog which arises from the abdominal wall and is inserted into the skin over the chest. The entire muscle contains only 3 or 4 muscle spindles, and

by the successive removal of portions of the muscle a preparation containing a single spindle with its nerve fiber intact was obtained. A sudden stretch (by means of a weight) applied to the muscle caused a rhythmical discharge of impulses along the nerve. The end organ apparently obeys the all or none law since increasing the strength of the stimulus above the threshold did not increase the *magnitude* of the electrical response. The *frequency* of the responses varied, however, directly with the intensity of the stimulus from 5 to 100 per second. A higher rate than this (300 per second) was obtained from a single touch receptor of the frog's skin by Adrian, Cattell and Hoagland, a blast of air being employed as the stimulus.

Matthews employed a small extensor muscle situated upon the outer side of the middle toe of the frog. This muscle contains a single muscle spindle. The afferent impulses were recorded by means of an amplifier and oscillograph from a small branch of the peroneal nerve supplying the muscle. The muscle spindle was stimulated by loading the muscle at various rates and with weights of different amounts. It was found that (a) the frequency of the impulses set up were roughly proportional to the logarithm of the load, (b) at 16°C the maximum impulse frequency obtainable was 290 per second. That is, each impulse followed its predecessor at an interval of only 0.0035 second. It follows therefore that each impulse traveled within the relative refractory period (0.01 second) of the preceding one. For this reason, some reduction (15 to 20 per cent) in the magnitude of the action potential waves occurs when they follow one another in rapid succession. Upon altering the temperature of the preparation the maximum frequency of the impulses varied in the same way as did the refractory period of nerve, that is, raising the temperature which has the effect of shortening the refractory period of nerve also shortens the interval between the impulses set up by maximal stimulation of the muscle spindle. (c) Passive stretch is an adequate stimulus to the muscle spindle provided the stretching force rises to its maximum at a sufficiently rapid rate. That the impulses recorded from the afferent nerve originated in the spindle was evidenced by the fact that when it was rendered functionless (by constricting it tightly with a thread) no impulses could be set up thereafter by stretching the muscle. It has been further shown by Matthews for mammalian muscle that though the two elements of the spindle ("flower spray" and annulo-spiral, p. 943) are stimulated by *passive* stretch, each behaves differently to *active* contraction of the muscle. The responses of the "flower spray" endings cease during contraction which apparently relieves the strain upon them. Those from the annulo-spiral endings also cease if the contraction is submaximal. With a supra-

maximal contraction, however, they increase in frequency, due then, it is believed, to a contraction of the intrafusal fibers. The Golgi endings in tendon respond to stretch whether passive or induced by contraction of the muscle.

Adaptation

The rhythmical discharge of impulses set up by stimulation of a receptor gradually diminishes in frequency and may soon cease entirely, though the stimulus continues to be applied at its original intensity. If, for example, the touch receptors of the cat's toe pad are excited by resting a glass disc lightly upon the skin a burst of impulses occurs at high frequency for from $\frac{1}{10}$ to $\frac{1}{2}$ second, but have entirely ceased by the end of $\frac{1}{2}$ second.

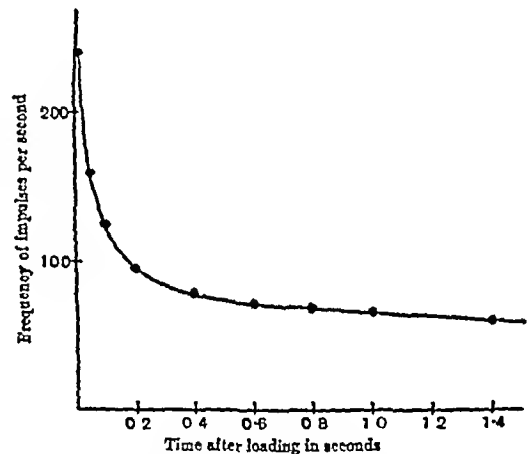


FIG. 63.7 Adaptation. Graph of response during first second after sudden loading with a large load 100 gm. Temperature 14°C (after Matthews).

After this, pressure exerted upon the disc is followed by a second series of impulses which again comes to an end though the pressure is maintained unchanged. This phenomenon is spoken of as *adaptation* (fig. 63.7). The gradual numbing of sensation to a constant stimulus is a common experience. We speak of "becoming accustomed" to some or other environmental change which initially caused a very definite effect. The olfactory endings, for instance, soon become adapted to the action of most odorous substances, and a light object placed upon the hand soon becomes imperceptible. We do not feel the contact of our clothes, the ticking of a clock after a time is not heard, and a hot bath feels much hotter when one first enters it than after a short period of immersion. The receptors of the retina also undergo

rapid adaptation, failing to respond after a brief period of stimulation. For this reason a small moving object is more readily perceived than a stationary one, and when one directs his attention to a motionless object the eyeballs are kept constantly moving in order to stimulate the receptors intermittently.

Adaptation should not be confused with fatigue. It occurs to the same extent whether the frequency of discharge is slow or rapid, if it were a fatigue phenomenon it should of course be more evident at rapid discharge rates. Furthermore, the excitability of the adapted end organ to a fresh stimulus is not reduced, and whereas fatigue is hastened by oxygen lack, adaptation is not. Adaptation is a property common to receptors whether of skin, muscle, viscera or special sense organs. Its rate of development, however, is different for the several types of end organ, being very rapid for those of touch, slower for those of pressure and warmth and still slower for the muscle spindle. As mentioned earlier, in the case of the bare nerve endings subserving pain, no true adaptation occurs. Receptors whose function it is to record sudden changes in the external environment adapt rapidly. This is necessary in order that, having responded to a given environmental change and conveyed their message to the central nervous system, the receptors shall again be ready to respond to another stimulus, and also that information received from receptors stimulated successively shall not be confused. We are aware for instance, of the movement of a light object over the skin since successive touch receptors respond instantly to the stimulus and adapt rapidly. If rapid adaptation did not occur, that is, if the discharge of impulses outlasted to any appreciable extent the duration of the external stimulus, the subsequent stimulation of neighboring receptors would result in a diffuse ill defined sensation. It is necessary, on the other hand, for the muscle spindle to adapt very slowly, since upon this receptor depends in part the reflex maintenance of a continued tonic contraction of postural muscles. It would seem requisite, on the other hand, that proprioceptors giving information concerning the *movements* of muscles and joints should adapt rapidly. There is little information, however, upon this point. The proprioceptors of the carotid sinus and aorta adapt slowly, those excited by inflation of the lung also have a slow rate of adaptation. Sudden forcible inflation of the lung, for example, causes a burst of impulses at the rate of 150 per second, by the end of the first second the frequency has only fallen to 115 per second.

A comparison of receptor and nerve fiber responses

In general the nerve fiber and the end organ (or rather the nervous element of the latter) behave similarly. Both require for excitation a stimulus

of a certain minimal strength and this must rise to its maximum value at a certain rate and have a certain minimal duration (p 946). Both obey the all or none law and have absolute refractory periods of about the same duration. There are certain differences, however, between the responses of the two structures, and some of these have already been pointed out. The sense organ possesses a low threshold for one type of stimulus whereas the nerve fiber responds unselectively to various types. The later stages of recovery (relative refractory period) of the end organ are slower than those of the nerve fiber. The sense organ also differs from the nerve fiber in giving a rhythmical response to persistent stimulation. When a constant current of moderate strength is

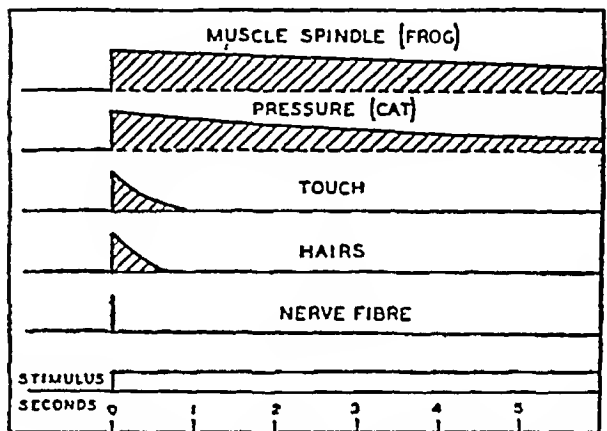


FIG 63.8 Showing the response of nerve fiber and of different types of receptor to a continued stimulus. Adaptation is most rapid in the nerve fiber and slowest in the muscle spindle (from Adrian, *The Basis of Sensation*, Christophers, London)

sent into a nerve a single impulse results, the refractory phase follows and though the current continues to flow a second response does not occur when the nerve recovers. If the stimulus, however, is very intense a short series of impulses may result. The apparent differences in response between the nerve fiber and end organ to persistent stimulation is therefore only quantitative. The nerve adapts rapidly, the sensory endings slowly. As a matter of fact the touch receptors adapt almost as rapidly as the nerve fiber itself (fig 63.8). The rhythmical nature of the end organ response is thought to be due to the stimulus causing a state of persistent depolarization in the surface film of the non-medullated ending within the receptor. The depolarized section of nerve then acts as a constant stimulus to the rest of the fiber. Lillie's iron wire model (p 926), for instance, will

respond rhythmically if the protective film of oxide is prevented from forming upon a terminal section of the wire. The maintenance of this part in a depolarized or active state causes a series of electrochemical reactions to be transmitted down the wire, the protective film forming upon the length of the wire after each wave has passed. Adrian also found that when the end of a nerve is injured rhythmical responses of high frequency are set up. These are probably the result of increased permeability of the damaged surface film, which therefore remains in a permanently depolarized state. These observations suggest that a mechanical stimulus applied to a sense organ causes deformation of its nerve ending with the production of analogous changes in the surface film which, persisting for a short space of time, cause the rhythmical discharge.

THE INTENSITY AND QUALITY OF SENSATION SPECIFIC NERVE "ENERGIES"

It has already been pointed out (p. 944) that the form and magnitude of the electrical response and so presumably of the nerve impulse are not altered by increasing the intensity of the stimulus applied to a receptor organ. So far as is known the only effect of raising the intensity of the stimulus is to increase the frequency of the impulses discharged. *Intensity of sensation* is believed, therefore, to depend upon the number of impulses reaching the sensorium per unit of time. The number of impulses reaching the sensory areas of the brain in a given time is dependent not only upon the frequency of the impulses arriving by each fiber but also upon the number of fibers involved. What may be termed the *massiveness* of the sensation is probably largely dependent upon the latter factor.

The relation between intensity of stimulus and impulse frequency may be accounted for upon the basis of the refractoriness of nerve. We have seen that the absolute refractory period puts an upper limit upon the frequency of the nerve impulses, and that a relative refractory period in which excitability increase gradually succeeds the absolute refractoriness of the nerve. It follows then that the stronger the stimulus the earlier in the relative refractory period, the shorter will be the intervals between the impulses, and, consequently, the higher will be the frequency of impulse production. All gradations of impulse frequency up to a maximum (when nerve responds at beginning of the relative refractory period) and down to a mini-

mum (nerve responding only at end of relative refractory period) by stimuli of various strengths. Thus different intensities of stimulation are signalled.

To what are due the different *qualities of sensation*? Is each type of sensation, e.g., touch, pain, sight, hearing, etc., mediated by a specific type of nerve fiber, that is, one with a distinctive type of end organ and which conveys impulses to a particular region of the brain? Is it for this reason that each type of end organ, in whatever way it is stimulated, always produces the same quality of sensation? On the other hand, is the type of sensation dependent upon the induction in non-specific nerve endings of some change characteristic of the type of stimulus applied, the characteristic effect of the stimulus being then conveyed as a message to consciousness? The first view, which is generally accepted, is spoken of as the *doctrine of specific nerve energies*. The principle was enunciated in 1838 by Johannes Müller, though it had previously been suggested by Sir Charles Bell of Edinburgh. It was extended by Helmholtz to apply to the different qualities embraced within a single type of sensation, as in his resonance theory of hearing—the appreciation of variations in pitch being dependent, it was claimed, upon the situation in the cochlea of the particular end organs which were stimulated. Also, in Young's theory of color vision the appreciation of the different primary colors was attributed to the presence of distinct sets of receptors.

The stimulus itself has no characteristic property, as was thought at one time, which is transmitted along the nerve fiber to consciousness, nor is the special nature of the sensation dependent essentially upon the receptor or the nerve fiber in the sense that it sends any characteristic message to the brain. The function of the receptor is to respond *most effectively* to one type of stimulus and to cause a discharge of impulses along the nerve fiber. The stimulus to which the receptor is designed to respond most effectively is called the adequate stimulus.

Some of these facts relating to so-called specific nervous energies⁵ can be recalled from common

⁵This term, coined by Müller in the middle of the 19th century, is rather confusing, but it means simply that each kind of peripheral sense ending always gives rise when stimulated to its own peculiar kind of sensation, and not that the several modalities of sensation are dependent upon types of nerve fiber differing from each other by some specific and distinctive form of activity.

experience The retina, for example, is stimulated most effectively by light waves (adequate stimulus) but it, or even the optic nerve itself, will also respond by a visual sensation to an electrical or mechanical stimulus, thus, flashes of light are experienced when the eye is struck and phosphenes appear when the eyeball is firmly pressed A mechanical stimulus, such as a blow, applied to the ear causes a sensation of sound The presence in the skin of definite "spots" for touch, pain, etc., also points to the different cutaneous sensations being dependent upon the adequate stimulation of specific endings, each of which transmits its message to the central nervous system along a definite pathway Just as the retina, however stimulated, gives rise to a visual sensation, so a touch spot arouses a tactile sensation, a warm or cold spot to a sensation of warmth or cold, and the stimulation of a pain spot to pain ⁶ The specific nature of the touch receptors in the skin of the frog is indicated by the experiments of Adrian, Cattell and Hoagland, these receptors cannot at any rate give rise to pain, for when they were stimulated to discharge impulses at the maximum rate (300 per second) there was no indication that the sensation was painful A stimulus which would cause an impulse discharge of this frequency from a pain ending would certainly cause severe pain Comparable results were obtained from the touch receptors at the roots of the hairs in the guinea pig It has also been shown that a pressure receptor (Pacinian corpuscle) cannot give rise to pain nor is it stimulated by temperature changes

Granted that specific fibers do exist, that is, fibers which, whatever the type of stimulus applied, give rise to a characteristic sensation, there is the further question,—are the messages emitted by the various end organs and transmitted along the nerve fibers identical, and do the qualities of sensation depend solely upon the connections which the respective fibers

eventually make with cells in the brain? Or, on the other hand, does the specific function of the fiber depend upon some distinctive character of the impulse which it transmits? For example, do the impulses set up by stimulating a pain ending differ in some significant way from those originating in a touch receptor?

Though in general the impulses, whatever their origin or whether motor or sensory, are closely similar, certain quantitative and qualitative differences do exist Erlanger and Gasser (p 925), for example, showed that the axon potentials of different fibers had different conduction rates and amplitudes Adrian and others have shown differences in the frequency of the impulses Impulses may therefore differ from one another in (a) conduction rate, (b) magnitude or (c) frequency ⁷

But there is no evidence that any of these different properties play a major role in determining the qualities of sensation Qualities of sensation are determined by the particular type of end organ stimulated (which is specially adapted to respond to one type of stimulus) and ultimately and essentially upon the situation of the nerve cells in the brain where the impulses arrive Electrical stimulation of the visual area of the cerebral cortex in man causes a visual sensation and of the auditory area a sensation of sound Or as someone has expressed it, if impulses arising in the retina were directed to the temporal lobe and auditory impulses to the occipital lobe we should hear the lightning and see the thunder

Weber-Fechner Law That the appreciation of changes in intensity of a stimulus does not proceed continuously and without interruption, but by a series of steps, was shown by Weber (1834), the subject was later treated mathematically by Fechner (1858) who postulated quite erroneously that the magnitude of a sensation is proportional to the logarithm of the stimulus intensity (Fechner's Law) The just perceptible change in the intensity of a given stimulus was found to be a constant fraction of the intensity of the preceding stimulus The fraction is not the same for different modalities of sensation Furthermore, it is constant only over a restricted intensity range This law is described further in ch 73

⁶ It was thought, not so long ago, that pain could be caused by the intense stimulation of any cutaneous sense organ, which, if true, would refute the doctrine of specific nerve energies

⁷ The impulses in the optic nerve during retinal stimulation have about the same frequency as those from a cutaneous receptor

Dr Sam Nath
M.B.B.S M.Sc (Med)
Reader in Physiology
Smt Medical College
Jaipur

CHAPTER 64

REFLEX ACTION

The synapse. The point of contact between neurons either in the central nervous system or in autonomic ganglia is called the *synapse*, a term introduced by Michael Foster in 1897.¹ The axon of the neuron divides into numerous filaments which terminate in small button like expansions known as *end feet* (*pieds* or *boutons terminaux*) which are applied to the surface of the body or dendrites of another neuron (fig 64 1). They occupy about 40 per cent of the surface of some nerve cells, e.g., those of the ventral horn of the spinal cord. A single

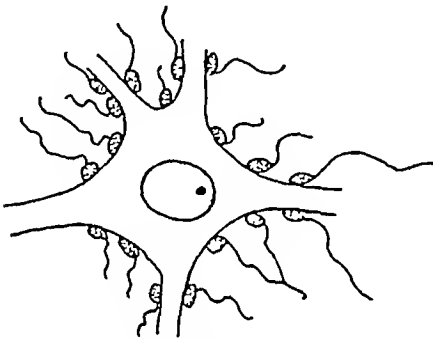


FIG 64 1 Diagram showing "end feet" (*pieds terminaux*) or synaptic knobs making contact with the body of a nerve cell

anterior horn cell shows several hundred (up to 2000) such contacts. It is generally taught that no structural continuity exists between neurons, but merely contact in the manner described. A passage of neurofibrils from an axon into the body or dendrites of another neuron can not be demonstrated. The fact that when a neuron is injured Wallerian degeneration (p 912) does not extend beyond its boundaries to involve a neighboring neuron also supports this conclusion. The end-feet themselves quickly show degenerative changes after the axon has been interrupted, within from 24 to 72 hours

¹ The word synapse as originally used referred to contact of axon with dendrite or with the cell body. Arvanitaki has introduced the term *ephapse* which applies to any contact between neurons.

² This does not apply to some invertebrates in which structural continuity between nerve cells does occur. But in such instances many of the features of reflex conduction are absent, which only strengthens the conviction that the presence of the synapse is responsible for the differences between reflex conduction and nerve fiber conduction.

they become swollen and at the end of about 120 hours undergo granulation and disintegration. The absence of protoplasmic continuity between two neurons implies that something in the nature of a surface film, though this be no more than a single molecular layer, is interposed. Such a film will provide a site for the development of surface phenomena, e.g., the formation of an electric double layer upon which the peculiar properties of the synapse probably depend. The synapse might therefore constitute a point of raised resistance—a higher threshold than that of the nerve fiber itself—which the impulse would need to cross or "step over" in order to reach another neuron. Or the synapse might provide a mechanism through which a *fresh* impulse is set up, rather than that it simply interposed new conditions in the path of the original one. Or, again, a chemical substance, e.g., acetylcholine might be released, which acted as a stimulus (through depolarization) to the second neuron. In the case of the central nervous system, neither of the last two alternative theories—electrical or chemical—has reached universal acceptance, both views have their adherents.

The functional connections between an efferent nerve fiber and the effector organ, e.g., the neuromuscular (myoneural) junctions where there is also structural discontinuity over which the impulse must be transmitted, are closely similar in their physiological nature to a synapse, particularly to those of autonomic ganglia.

THE CHARACTERISTIC FEATURES OF REFLEX ACTION

Spinal reflex action has been studied intensively by Sherrington and his associates, a number of phenomena have been discovered which cannot be demonstrated in the nerve fiber itself. In the following paragraphs (after Sherrington) the main differences between reflex conduction and nerve conduction are listed and briefly discussed.

(1) *One way conduction*—centripetal in the afferent and centrifugal in the efferent limb—as compared with two-way conduction in nerve fibers. The synapse may be compared to a valve or a gate which opens only one way. The irreversibility of

conduction in the reflex arc was demonstrated in the early part of the 19th century by Bell and Magendie. The former (1811) showed that stimulation of the anterior roots causes muscular contraction, but no motor effect followed stimulation of the posterior roots. Magendie (1822) found that stimulation of the posterior roots caused pain. These observations have been embodied in the so-called Bell-Magendie law which states that the anterior root fibers are purely efferent, those of the posterior roots purely afferent.

One-way conduction is not, however, an unalterable property of the synapse, for it is abolished by lowering the temperature of the spinal cord by a few degrees. It has been shown by Toennies, for example, that if the cord is cooled, stimulation of a dorsal root causes a discharge of impulses from the center which travel peripherally over a certain proportion of the fibers composing other dorsal roots, that is, antidromically (p. 276). Barron and Matthews have also found that cooling the cord causes muscular twitchings, and that each twitch is accompanied by a discharge of impulses along sensory root fibers.

(2) *Slower speed of conduction*, as measured by the latent period between the application of the stimulus to the afferent limb of the reflex arc and the response of the effector organ (e.g., contraction of a muscle) (p. 951).

(3) *Less close correspondence between the grading of the stimulus and the reflex response*. Reflexes vary greatly in this respect. When stimuli of gradually increasing intensity are applied in succession to a motor nerve the muscle shows steplike increases in response as a result of more and more nerve fibers being excited and of an increasing number of muscular units being brought into play, but after this an increase in the strength of the stimulus causes no greater effect ("all or none" principle). In the case of the reflex response, increasing the strength of a single stimulus beyond that which brings all the afferent fibers into play may cause a further increase in the response of the muscle. On the contrary, a reflex such as the extensor thrust, the strength of the stimulus and the response may show no correspondence, the latter may be "all or none" being maximal with either a weak or a strong stimulus. In other reflexes, such as the scratch reflex, grading can be obtained without difficulty, the amplitude of the scratching movements, though not the frequency, becoming greater with increasing intensity of stimulus. In those re-

flexes which can be graded in this way, the grading is probably due to an increasing number of motoneurons being excited by the stronger stimulation as a result of the spread or irradiation of the excitatory state within the spinal cord.

† (4) As a result of a repetitive discharge from the center ("after discharge," p. 952) there is a lack of correspondence between (a) the moment of cessation of the stimulus and the termination of the response of the effector organ, and (b) the rhythm of the stimulus and that of the muscular response.

(5) Considerable resistance in the reflex arc to the passage of a single impulse, but the ready forcing of the resistance by a series of impulses.

(6) *Greater variability in the threshold value of the stimulus in the case of the reflex arc*. The response to a given stimulus is influenced by conditions which do not come into play in the case of nerve fiber conduction, e.g., facilitation, recruitment, occlusion, etc. These will be considered later.

(7) *Susceptibility of reflex conduction to fatigue*, whereas nerve trunks are almost indefatigable.

(8) *Much greater susceptibility of reflex conduction to oxygen lack, anesthetic and other drugs, shock, etc*. This is evidently due to the presence of the nerve cell in the conducting pathway. Nerve cells have a much higher metabolism than nerve fibers (p. 930) and reflex excitability is abolished if the centers are deprived of oxygen for a short period. The cells of the cerebral cortex are irreparably damaged if deprived of their blood supply for 5 minutes or so, in man deprivation of blood for even half a minute may cause injury. The centers of the brain stem survive for a longer period—from 25 to 35 minutes, and the spinal centers for from 40 to 60 minutes. The autonomic centers are also relatively resistant to oxygen lack. According to Kabat, the resistance of the brain in very young animals (puppies up to 4 months of age) to circulatory arrest is much greater than is the adult brain. In deep *anesthesia* many reflexes are completely abolished, whereas nerve trunks remain excitable and may even respond to stimulation after an animal has been killed by an overdose of anesthetic. Ether vapor or other anesthetics applied *directly* to the nerve, however, quickly abolish its excitability.

Some of the special features of reflex conduction mentioned above will be dealt with in more detail in the following pages. Certain other characteristic features of reflex action will also be considered.

SITUATION THE CENTRAL EXCITATORY STATE

If a single weak stimulus which when applied to a motor nerve causes contraction of the muscle, is applied to the afferent nerve it usually fails to cause a reflex response. However, if the afferent nerve is stimulated by two or more stimuli in rapid succession, each of which is subliminal, a reflex contraction of the muscle follows. This summation of stimuli is well illustrated by the *scratch reflex*, as shown by the spinal dog. It consists of a series of rhythmical scratching movements of the hind-limb when the skin within a saddle-shaped area of the back is stimulated. It cannot be elicited by the application of a single induction shock or even by two shocks unless they are very intense. The reflex is readily elicited by a series of rapidly repeated weak stimuli. Sometimes the response does not occur until after the fortieth repetition (Sherrington). A single *medial* stimulus applied to the skin (e.g., plucking a hair, a pinch, etc.) may elicit the reflex, but under such circumstances it is probable that the receptor organ discharges a series of impulses (see p. 945).

The elicitation of the reflex by a rapid succession of subliminal stimuli is taken to mean that an excitatory state has been created in the reflex center by each subliminal stimulus, and which, through addition or summation, is built up to threshold value, a discharge of impulses then being discharged down the motoneuron. This central excitatory state (*c.e.s.*) decays rapidly lasting at any single synapse for from 0.5 to 2 msec. only. The summation just described as occurring in the scratch reflex is not true *temporal* summation, that is, summation at any individual synapse, but is really an example of *spatial* summation—summation of excitatory states of several synapses (see next section). In order for temporal summation to occur, subliminal stimuli would need to be separated by intervals longer than the refractory period of the presynaptic afferent fiber. But decay of the *c.e.s.* reaches a very low value within this period. Lorente de N6 found that in experiments in which the oculomotor nucleus was directly stimulated, summation was insignificant when the intervals between stimuli were greater than 0.1 to 0.2 msec., which is shorter than the refractory period of the afferent fiber. However, since the impulses of a volley² discharged along an afferent nerve traverses internuncial neurons in their

course to the motoneurons the excitatory state at the center persists for a much longer time (up to 20 msec.) Succeeding volleys reaching the center during this time will raise the *c.e.s.* to threshold value. But this, as just indicated, since several synapses are involved is actually a spatial summation. Temporal summation is of little physiological importance in the central nervous system.

Spatial summation—If two subthreshold stimuli are applied simultaneously, one to each of two afferent nerves, summation occurs and a response results. This type of summation, which is of essential importance in central excitation, is spoken of as *immediate (direct) spiral induction* or *spatial summation*. Allied reflexes mutually support or reinforce one another through such an effect. This therefore constitutes one type of *facilitation*. Immediate induction can be readily demonstrated in the scratch reflex (rhythmical, alternate flexion and extension at hip, knee and ankle). When two subthreshold stimuli are applied within the receptive area of the reflex but about 10 cm. or less apart a response follows. Each of the two scratch reflexes which would be evoked by a single stimulus though closely allied are not identical. The further apart the stimulated points the less alike are the respective reflexes, and the less pronounced is the mutually facilitating effect. Such a result, taken in conjunction with other observations which will be considered presently, indicate that the two afferent pathways converge upon and make connections with the same motoneurons.

REFRACTORY PERIODS

The nervous centers show an absolute and a relative refractory period, as well as a phase of supernormal excitability. The length of the absolute refractory period as observed by Eccles for the motoneurons of the flexion reflex is 2.5 msec. and from 10 to 15 msec. for the period of relative refractoriness. A lower value, namely 0.1 msec., was obtained by Lorente de N6 for the motoneurons of the more rapidly acting eye muscles. The refractory period is determined by applying a strong stimulus to the undivided motor nerve of the reflex, and thus sending a volley of impulses backwards, that is, antidromically, into the spinal center. The antidromic volley causes discharge of the center just as though it had been stimulated through the afferent nerve. The succeeding interval during which the reflex cannot be elicited by stimulating the afferent nerve is the absolute refractory period. This is followed by a period

² This term means a synchronous discharge of impulses in a number of nerve fibers, like the firing together of rifles by a squad or platoon of soldiers.

of subnormal excitability—the relative refractory period

REFLEX CONDUCTION TIME

The time elapsing between the stimulation of the afferent nerve or receptor and the response of the effector organ is called the *total latent period* of the reflex. If from this be subtracted the time which the impulse would take in traveling over the afferent and efferent fibers (which conduct at rates up to 120 meters per second) the time taken by the impulse to pass through the cord is obtained.⁴ This is called the *central or reduced reflex time*. In the table given below the total latent period is 10.4 m sec. and the total time of the peripheral pathways is 6.5 m sec. leaving 3.9 m sec. as the central reflex time. Since the length of the central pathway is only a small fraction of the length of the peripheral path while the central reflex time is nearly $\frac{1}{3}$ of the total time, it is clear that the conduction rate through the center must be relatively very slow. The delay presumably occurs at the synapse and is attributed to the time required to build up the excitatory state to threshold value.

Total latent period	10.4 m sec
Time for passage of impulses in 13.8 cm of afferent nerve at 31.6 meters per second	4.4 m sec
Time for passage of impulses in 19.5 cm of efferent nerve at 93 meters per second	2.1 m sec
Total time for peripheral path	6.5 m sec.
Therefore central reflex time (synaptic delay)	3.9 m sec.
—Modified from Creed, Brown, Eccles, Liddell and Sherrington	

When the interval between two stimuli applied to an afferent nerve is sufficiently brief the latent period of the response caused by the second stimulus is shorter than usual. The shorter latency cannot be attributed to a reduced conduction time in the peripheral nerve fibers, it must be central in origin, i.e., at the synapse. Now, the central excitatory state rises to a maximum during a measurable time and declines again gradually. The reduced latency of the second volley of impulses is therefore attributed to their reaching the

synapse before the effect of the first volley has disappeared, therefore a shorter time is necessary for building up the C.E.S. to threshold value. The synaptic delay under these circumstances may be reduced to 0.5 m sec.

The delay is also greatly reduced by increasing the strength of the stimulus and may then be so slight that the conduction rate over the arc as a whole approaches that over nerve trunks.

SUCCESSIVE INDUCTION

This term is applied to the facilitation which one reflex exerts upon another immediately succeeding it (immediate or direct induction is described on p. 950). Successive induction may be *positive* or *negative*, that is, the second reflex reaction may be of the same sign as and allied to the first or of opposite sign and antagonistic. An example of positive successive induction is the augmentation of the scratch reflex which follows the successive stimulation of two separated points in the skin, as by rolling a spurred wheel over the surface. A parasite moving over the skin will exert a similar effect. Perhaps, as Sherrington suggests, the long hop of the flea has evolved as a means of circumventing facilitation of this reflex. Negative spinal induction is seen in rhythmical or alternating reflexes such as stepping, the shake reflex of the head or the biting reflex. For instance, the flexion reflex (contraction of flexors and relaxation of extensors of the limb) favors a succeeding extensor reflex (contraction of extensors and relaxation of flexors) which in turn lowers the threshold for the next flexion reflex.

In the spinal animal (p. 967) spontaneous automatic stepping or scratching movements may occur which under favorable conditions persist for several minutes. Automatic stepping movements of the hind limb of the cat may persist even under deep anesthesia and after the opposite half of the cord of the segment supplying the limb has been destroyed. The reflex alternating movements are not abolished by deafferentation of the limb (section of posterior roots). The perpetuation of the rhythm is therefore not due to impulses arising in the contracting muscles, it is central (spinal) in origin.

REBOUND

Allied to negative spinal induction is the phenomenon known as *rebound*. When, for instance, during the elicitation of a reflex, inhibition is

⁴ If the time of arrival of the action current at the end of the efferent nerve is recorded the latency of the muscular response need not be considered.

induced by the stimulation of an afferent nerve, augmentation of the reflex occurs when the inhibitory stimulus is withdrawn (fig 64.2) The knee jerk when inhibited by stimulation of the central end of the hamstring nerve is more brisk after cessation of the inhibition The rebound has a long latent period occurring about 0.2 second after the withdrawal of the inhibition Sometimes the mere cessation of the stimulus which excited the reflex is sufficient, without inhibition, to cause rebound It is most readily elicited in the crossed extension reflex of a decerebrate preparation, but is elicited with difficulty in the spinal animal Rebound is easily inhibited It appears after the silent period of a tendon jerk (p 961) and, under certain favor-

induction apparently act together to increase the force of a contraction following a phase of inhibition For example, electrical or mechanical stimulation of the gums or palate of a decerebrate preparation causes inhibition of the closers of the jaws (temporals and masseters), and contraction of the jaw openers (digastrics) Immediately upon cessation of the stimulus, the jaw-closers contract powerfully Thus, a rhythmic chewing reflex tends to keep itself going, so long as there is any material between the jaws upon which to bite (Sherrington)

AFTER DISCHARGE

The discharge of impulses from the reflex center and the consequent prolongation of the reflex response after the stimulation of the afferent nerve or receptor has ceased is called *after discharge* (fig 64.3) For example, the crossed extension reflex of the dog (extension of knee, ankle and hip following stimulation of the afferent nerve of the contralateral limb) may have, if the stimulus is very intense, an after discharge of a

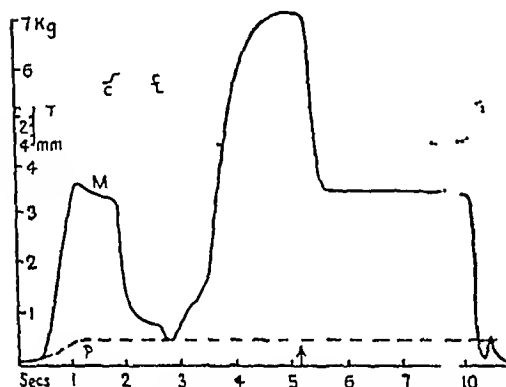


FIG 64.2 Between c and c' the stretch reflex of vastus medialis is inhibited by faradization of the peroneopopliteal nerve of the same side Note the inhibition and the further fall, "post inhibitory notch" after c' The post inhibitory notch is followed by considerable rebound which in turn is inhibited at } by gentle traction on biceps femoris of the same side (from Creed, Denny Brown, Eccles, Liddell and Sherrington, *Reflex Activity of the Spinal Cord*, The Clarendon Press)



FIG 64.3 Flexion reflex, showing *after discharge*. The duration of the stimulus is indicated by the pair of nearly vertical lines (after Sherrington)

second or more After discharge of the flexion reflex may last for as long as 8 seconds In other reflexes, e.g., the flexion and scratch reflexes, the after discharge is much shorter than this (60 to 70 msec) After discharge is spoken of by Sherrington in a figurative sense as reflex "momentum" Increase in the strength or duration of the stimulus lengthens the after discharge Increase in the intensity of the stimulus has a relatively greater effect upon the duration of after discharge than upon the amplitude of the contraction, after the latter has been increased to the maximum further strengthening of the stimulus lengthens the after discharge After discharge is readily inhibited

As a result of the after discharge from the spinal center a single stimulus applied to an afferent nerve is followed, not by a single twitch as when the motor nerve is stimulated, but by a series of twitches, an incomplete tetanus or even a complete tetanus When a series of stimuli at a rate under 50 per second are sent into an afferent nerve

able conditions, is followed in turn by inhibition and this again by rebound Such a series of inhibitions and rebounds following the sudden stretching of a tendon and giving rise to automatic jerking movements is spoken of as *clonus* The cause of rebound is not known It is not due to proprioceptive impulses, since deafferentation of the muscle does not abolish it

One explanation which has been offered is, that when the inhibitory stimulus is applied, excitatory afferent fibers as well are stimulated but their effects are masked by the inhibitory effect. After withdrawal of the stimulus the persistence of the "after discharge" of excitation beyond the inhibitory "after action" accounts for the augmented contraction In certain rhythmical reflexes such as biting, rebound and negative successive

the rhythm is just perceptible in the myogram, whereas in the tetanus caused by direct stimulation of the motor nerve at the same rate the individual contractions are clearly visible. The smoothing out of reflex rhythm is due to the after discharge of impulses following each main volley, the tension of the muscle between the latter is thereby sustained.

After discharge is the result of the persistence and gradual subsidence of the excitatory state built up in the reflex center by the afferent nerve stimulation and the consequent repetitive discharge of impulses along the efferent fibers. The most probable explanation of the prolonged discharge is the one proposed by Forbes, namely, that impulses are set up in certain long internuncial circuits ("long circuiting") and owing to the length of the course over which they must travel are delayed in their arrival at the motoneurons. It is unlikely that the asynchronous arrival of impulses at the center, due to different conduction rates of the fibers composing the afferent nerve or of variable delays at synaptic junctions, plays any important part, the after discharge is too long to be accounted for in these ways. The arrival at the center of impulses set up by the contracting muscle itself is of minor importance. If the afferent impulses from the muscle are blocked by sectioning the posterior roots (deafferentation) the after discharge is reduced though not abolished, it is therefore dependent, in part only, upon the reexcitation of the center by proprioceptive impulses.

RECRUITMENT

Many reflexes gradually increase to a maximum when a stimulus of *unaltered* intensity is merely prolonged. This is due to the activation of a progressively greater number of motoneurons. The phenomenon is called *recruitment* and is figuratively spoken of as "*inertia*" by Sherrington. As already mentioned, "after discharge" is referred to as reflex "momentum." Both features, neither of which is in evidence when the motor nerve is stimulated directly, obviously must give smoothness to reflex action. Recruitment is more easily demonstrated in decerebrate than in spinal preparations, it is well seen in the crossed extension reflex but not in the flexion reflex (p. 957). It is also seen in inhibitory reflexes, i.e., prolongation of inhibitory stimulus of constant strength causes the inactivation of an increasing number of motor neurons.

IRRADIATION OF REFLEXES

When, in a decerebrate animal, the strength of a stimulus applied to a receptive area is gradually increased the central excitatory process spreads to a progressively greater number of motoneurons. As a consequence, additional muscle groups take part in the reflex response. The phenomenon is called *irradiation*. The impulses do not spread indiscriminately to all the motoneurons within a given radius of the point of entry of the afferent fibers, on the contrary, owing to the variability of the resistance at different synapses they are directed along selected paths. The synapses at the motoneurons in close proximity to the afferent terminals have the lowest threshold, but as the strength of stimulus is increased impulses spread further afield and excite the neurons of adjacent and ultimately of more remote segments. For example, a weak

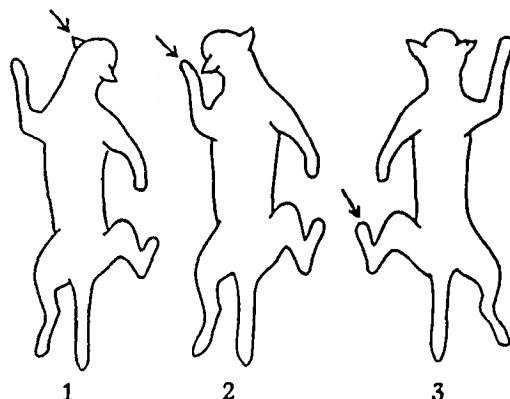


FIG. 64.4 Reflex figure. Arrows indicate the parts stimulated.

stimulus applied to the sole of the hind paw of a decerebrate animal causes flexion of the ankle alone, upon gradually increasing the strength of the stimulus the knee flexes, then the hip (flexion reflex), later, extension of the crossed hind limb occurs (crossed extension reflex), extension of the ipsilateral forelimb follows and finally flexion of the contralateral forelimb.

Besides the soles there are six other areas from which irradiation can be demonstrated very easily. These are the palms, the pinnae, the snout and the tail. The final attitude assumed when a stimulus of suitable strength is applied to the sole or to one of the other reflexigenous areas just mentioned is spoken of as a *reflex figure* (see fig. 64.4). The spread of the irradiation (march of the reflex figure) in all cases follows a definite order. Upon stimulation of the pinna the order of spread is — lateral flexion of neck to the opposite side, flexion

of ipsilateral forelimb, extension of ipsilateral hindlimb and extension of contralateral forelimb. The order upon stimulation of the forelimb paw is—flexion of stimulated forelimb itself, extension of ipsilateral hindlimb, flexion of contralateral hindlimb, extension of contralateral forelimb. Corresponding reactions are produced by stimulation of a hindlimb paw—flexion of the stimulated limb, extension of opposite hindlimb and tail, extension of ipsilateral forelimb, and flexion of contralateral forelimb.

It will be noted that only those muscles which act synergistically, that is, those which combine to effect a purposeful coordinated act, are activated. Muscles whose action is antagonistic to these are inhibited. The reaction which results corresponds to that which would occur from stimulation of the intact animal. Thus a painful stimulus of the hind paw causes withdrawal of the injured part, the other limbs reacting at the same time to support the body and move it from the stimulating agent. The reflex response is therefore quite different from the response which would result from stimulation of an anterior nerve root—all muscles to which the fibers of the root were distributed would, of course, be activated.

FRACTIONATION OF THE REFLEX CENTER

When an afferent nerve is stimulated by a faradic current of *gradually increasing* intensity the tension developed by the reflex contraction increases to a point after which increase in the strength of the stimulus fails to augment the reflex tension. The limitation is not simply due to all the motoneurons of the reflex having been excited, for direct stimulation of the motor nerve causes a development of tension greatly in excess of that resulting from the stimulation of the single afferent nerve. *It is concluded that under the ordinary conditions attending reflex action, a single afferent nerve when stimulated activates only a fraction of the motoneurons supplying a given muscle.* The size of the fraction of the motoneurons governed by the afferent nerve varies under different conditions, it is reduced by fatigue, oxygen lack, anesthesia, etc., and is increased by the administration of a small dose of strychnine. Even under the most favorable conditions, however, the number of motoneurons excited through a single afferent nerve is not more than a fraction of those which can be excited by stimulating the motor nerve.

Furthermore, the motor units⁶ excited through one afferent nerve do not lie in a single muscle, but are distributed throughout many different and often widely separated muscles. In table 90 are shown the reflex tensions developed respectively in the tibialis anticus and semitendinosus as a result of stimulation of various afferent nerves. In the last column these tensions are given as percentages of those developed by direct stimulation of the motor nerves (motor tension). It will be noted that the fraction of the motor units activated by the different afferent nerves varies

TABLE 90

MUSCLE	AFFERENT NERVE STIMULATED	MAXIMUM REFLEX TENSION	MAXIMUM REFLEX TENSION IN PER CENT OF MAXIMUM MOTOR TENSION
		GRAMS	
Tibialis anticus	Internal Saphenous	800	32
	Ext. Cut. of groin	1,060	37
	Superf. obturator	165	6.7
	Deep obturator	400	16
	N. to quadriceps and Sartorius	1,190	44
	Small sciatic	1,320	45
	Popliteal	1,820	78
Semitendinosus	Internal Saphenous	1,900	63
	Ext. Cut. of groin	830	28
	Superf. obturator	1,270	42
	N. to quadriceps and Sartorius	2,740	91
	Small sciatic	?	62

widely, those nerves more closely allied segmentally to the motor innervation involving a larger

⁶ By a "motor unit" is meant a single motor neuron and the bunch or "squad" of about 100 muscle fibers which it supplies. Close to the muscle and within it, each nerve fiber divides dichotomously into several daughter fibers, each of which ends in one of the 100 or so muscle cells. The tension developed by a single motor unit was determined by dividing the total number of motor units in a muscle (arrived at by actual count of the fibers before branching in a motor nerve) into the tension developed when the motor nerve was maximally stimulated. It was found that a single motor unit of the cat's gastrocnemius developed a tension of 30 grams. The semitendinosus muscle of the cat has some 630 such motor units, the gastrocnemius medius 430, the extensor longus digit 330 and the soleus 250. The values found by Eccles and Sherrington for the tensions developed were, per motor unit, 5.5, 30.1, 9.9, 8.6 grams respectively for the semitendinosus, gastrocnemius medius, soleus and extensor longus digit.

proportion of the units of the muscle than those more remote

These facts are important in understanding the distribution of paralysis involving the anterior horn cells (e.g., anterior poliomyelitis). The pattern of the paralysis is according to motor units rather than to whole muscles. Thus a discrete group or fasciculus of degenerated muscle fibers may be found surrounded by normal muscle.

THE PRINCIPLE OF CONVERGENCE, OCCLUSION

Evidence that a given group of anterior horn cells is in communication with the central terminals of more than one afferent nerve has already been cited in the paragraphs dealing with spinal induction. This principle of convergence of afferent pathways is shown in other ways. For example, if a muscle be strongly excited by the successive stimulation of two afferent nerves it is found that the sum of the reflex tensions developed in the two responses is much *greater* than the maximal tension developed after a single stimulation of the *motor* nerve. Obviously there must be a central overlap of the two afferent pathways—a proportion of the motoneurons supplying the muscle must be connected with both.

Also, the sum of the reflex tensions developed when strong stimuli are applied to two afferent nerves successively is often found to be considerably greater than the tension resulting from stimulation of the two nerves simultaneously. The difference in the tension values is called *occlusion*. For instance, the sum of the tensions developed in the tibialis anticus when the two plantar nerves (afferent) are stimulated *consecutively* is 3.15 kg (1.57 + 1.58 kg) whereas the tension following *simultaneous* stimulation of the same nerves is only 1.81 kg (no more than 15 per cent greater than that resulting from the stimulation of one nerve). The occlusion therefore amounts to $(3.15 - 1.81) = 1.34$ kg. This result is not due to interference of one afferent pathway with the other, or to inhibition. It is simply due to the fact that a certain proportion of the motoneurons of the reflex is common to the two afferent nerves, the motoneurons common to the two afferents are excited maximally by either nerve alone, so no greater effect upon the muscle can be produced by exciting them through both afferents simultaneously. Occlusion is therefore a measure of central overlap, i.e., of convergence of afferent paths (fig. 64.5 A).

Not only in reflexes set up through afferent nerves of similar function and distribution does occlusion occur, it is also evident between reflexes resulting from

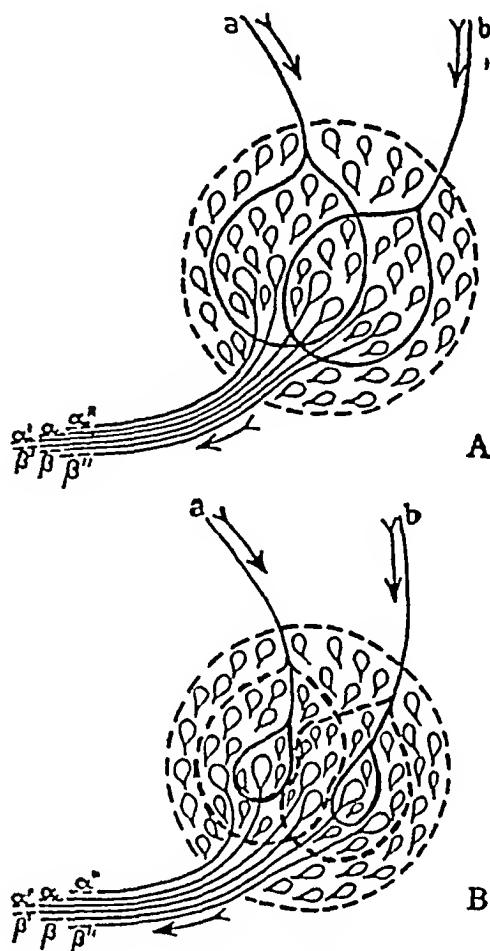


FIG. 64.5 A, illustrating occlusion. *a* and *b* are two afferent fibers, *a* when stimulated alone excites the motoneurons within the outlined area, only four motor fibers (α' , α , α' and β') are represented. *b* when stimulated alone activates the motoneurons in the corresponding area, of which four axons (β' , β , β'' and α') are shown. Thus when the two afferents are stimulated consecutively the motoneurons α' and β' are activated each time, making a total of eight. When stimulated concurrently the total number of activated motoneurons is six, i.e., there is a deficit of contraction due to occlusion. B, weaker stimulation of the afferents *a* and *b*, restricts the respective fields of threshold excitation as shown by the continuous line limit. *a* by itself activates one motoneuron α , *b* activates β . Concurrently they activate four units (α' , α , β' and β) owing to the summation of the subliminal effect in the overlap of the subliminal fields outlined by the broken lines (from Creed, Denny-Brown, Eccles, Liddell and Sherrington, *Reflex Activity of the Spinal Cord*).

stimulation of such dissimilar nerves as the internal saphenous and the nerve to the sartorius. The smoothness of reflex action, the interlocking of allied reflexes, the gradual, even passage from one reflex to another

and the general coördination of muscular function are the natural consequences of, (1) the fractional control by any one afferent of the centers governing the muscles of a given reflex, and (2) the overlap of the motoneuron fields of afferents governing different reflexes

"**SUBLIMINAL FRINGE**" Increasing the strength of the stimulus increases the extent of the central overlap and so of the occlusion. When, on the other hand, the stimulation is weak, occlusion may be absent and the sum of the tensions resulting from the concurrent stimulation of the two afferents will then be greater than the tension developed when they are stimulated successively, i.e., instead of there being a deficit of contraction there is actually augmentation (facilitation). The reason for this will be made clear by reference to figure 64.5. Strong stimuli set up effective excitatory states in the motoneurons connected with each afferent path, these excitatory fields overlap and concurrent stimulation of the two afferents results in occlusion as just described. Fringing the effective excitatory field (*liminal field*) is a zone in which the excitatory state is beneath the threshold for the discharge of impulses down the motoneurons. This is called the *subliminal fringe* (it is not shown in fig. 64.5 A). When the stimuli are weak the liminal fields are less extensive and so do not overlap (fig. 64.5 B). Overlapping of the *subliminal fringes* of the two afferents, however, causes through summation the excitatory state of the overlapped neurons to be raised to the threshold level. Thus motoneurons which are unexcited by either stimulus alone, are excited when both are applied together. The response which results when subthreshold stimuli are applied simultaneously to two afferent nerves (immediate spinal induction) may also be explained by the overlap of subliminal fields.

That subliminal fringes surround excitatory states produced by *maximal* as well as those produced by weak stimulation is shown by the following experiment. Strong faradization of the central end of the internal saphenous nerve causes tetanic contraction of the tensor fasciae femoris. A single break shock applied to the musculo-cutaneous nerve induces a reflex contraction which lasts for $\frac{1}{2}$ second. Stimulation of the musculo-cutaneous *during* the tetanic contraction induced by the stimulation of the internal saphenous causes an augmentation of the contraction which persists for $\frac{1}{2}$ second. The facilitation is attributed to the subliminal fringe caused by stimulation of the musculo-cutaneous being overlapped, and so raised to threshold value, by the subliminal fringe caused by stimulation of the internal saphenous.

The subliminal fringe provides a mechanism for the linking together of allied reflexes having origins close together or widely separated. It provides a background or, as Sherrington expresses it, a "catch on" for labyrinthine and other types of reinforcement. Stimulation of certain areas of the cerebellum, for instance, produces no effect itself upon limb movements, yet can modify reflex movements already in progress. Again, afferent stimuli too weak to elicit a limb reflex become effective when neck and labyrinthine proprioceptors are stimulated by turning the head, though this movement itself causes no apparent effect. The facilitation of one phase of an alternating reflex by the other (p. 951) is also attributed to the overlap of subliminal fringes persisting for a short period after each phase.

THE PRINCIPLE OF THE FINAL COMMON PATH

Each fiber composing an afferent nerve is a pathway for impulses arising in a limited number of receptor organs (ch. 64). The afferent nerves are private paths through which the receptors communicate with the spinal centers. The motoneurons of a given reflex, on the other hand, are as public roads which must accommodate traffic from a large number of differently located points, upon them impulses from many and various receptive areas of the body converge. This final link in the reflex arc, the motoneuron, is therefore called the *final common path*. Each final common path (F.C.P.) is as the stem of a funnel of which the numerous afferent paths represent the expanded portion (see fig. 63.2, p. 934 and fig. 64.6). Conversely, each afferent nerve can communicate with a great many F.C.P.'s. Potentially, every afferent fiber is in communication with all F.C.P.'s. After strychninization, for instance, or the injection of tetanus toxin, a stimulus applied to practically any region of the body induces wide spread muscular contractions (convulsions). Since all reflex action of muscle is brought about through impulses traversing the final common path, this must subserve a variety of reflex responses. At the entrance to the final path, i.e., at the convergence point of the afferent pathways, excitatory and inhibitory states are set up. It is upon the nature and extent of these that the type and intensity of the resulting reflex depend.

Allied reflexes, i.e., those conjoining to effect a common purpose, use the common path concurrently and reinforce one another. Such reflexes may originate in widely separated regions and from

diverse receptors (receptors of skin, eyes, organs of smell, muscle and labyrinthine proprioceptors, etc.) For example, stimulation of the outer digit of the hind foot causes reflex of the same leg. If other digits are stimulated concurrently the flexion reflex occurs more readily (facilitation, p. 950). Additional stimulation of the contralateral fore foot or pinna or of the tail further reinforces the flexion response.

Antagonistic reflexes, on the contrary, cannot occupy the final path simultaneously, the less important of two such reflexes must give way. If, for example, during the scratch reflex a strong stimulus is applied to an afferent nerve of the limb engaged in the scratching movements, the limb becomes strongly flexed, i.e., the scratch reflex

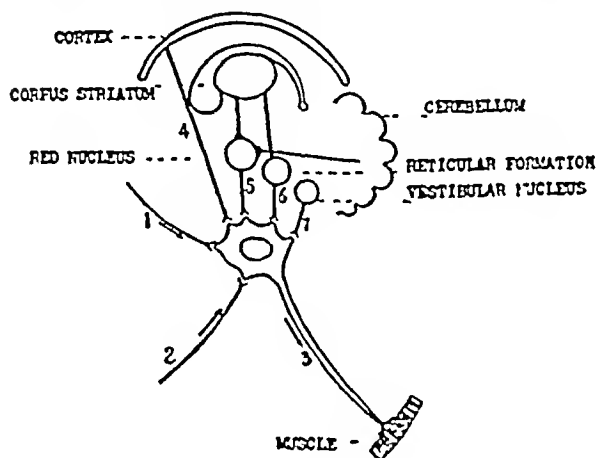


FIG 646 Diagram showing the convergence of various paths upon the anterior horn cell 1 and 2, afferent fibers from periphery, 3, motoneuron (final common path), 4, corticospinal fiber, 5, rubrospinal fiber, 6, reticulospinal fiber, 7, vestibulospinal fiber. The pathway from the tectum to the spinal motoneuron has been omitted. (In part from Lennox and Cobb, redrawn and modified)

gives place to the flexion reflex. Protective reflexes, e.g., those set up by injurious agents and accompanied by pain in the conscious animal are prepotent, displacing less urgent responses from the final common path. The flexion reflex (see below) which withdraws the foot from a noxious influence is a reflex of this nature.

RECIPROCAL INNERVATION INHIBITION

Voluntary or reflex contraction of a muscle, e.g., the biceps, is accompanied by the simultaneous relaxation of its antagonist—the triceps. Reciprocal innervation of antagonistic muscles is illustrated by the following experiment. When the paw of a spinal or decerebrate animal (p. 964) is pinched, pricked, burned or stimulated by a

strong electric shock, or the central end of the peroneal nerve excited, there results in the same limb a strong contraction of the flexor muscles of the ankle, knee and hip, together with inhibition of the extensors. This is the *flexion* or *withdrawal* reflex. In the contralateral limb the picture is reversed. There occurs contraction of the extensors together with inhibition of the flexors—the *crossed extension reflex* (fig. 647 and 648). The purpose in these reflex reactions is obvious, when an animal's paw is painfully stimulated the

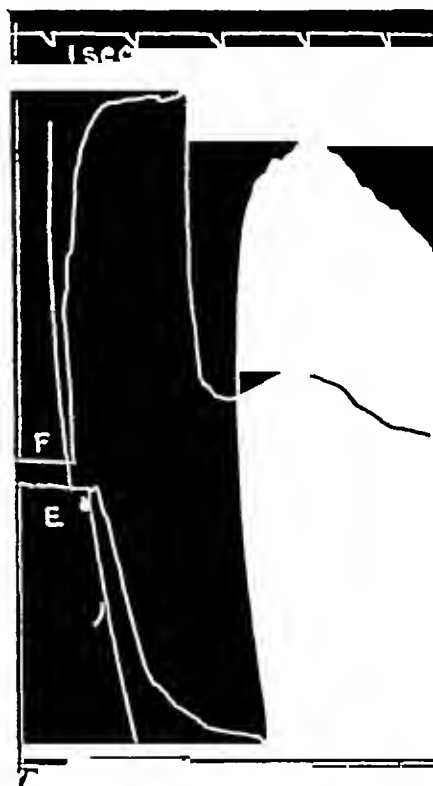


FIG 647 Reciprocal action of antagonistic muscles. Record from leg of a decerebrate cat, F, flexor, E, extensor. The myograph writer for the extensor muscle is set a little to the right of that for the flexor muscle in order that the two may clear each other. The ascent of F and the descent of E are actually synchronous (from Sherrington)

flexor muscles contract to withdraw the limb while contraction of the extensors of the opposite limb stiffen it for the support of the body.

After poisoning with strychnine reciprocal inhibition is abolished. Both sets of muscles, agonists and antagonists, respond to stimulation by simultaneous contraction.

The reciprocal inhibition of the knee extensors in the flexion reflex may be demonstrated more clearly after abolishing the action of the flexor muscles by sectioning their tendons, stimulation of the afferent nerve is followed by relaxation of

the extensors, and the limb, if it be in such a position that it can be acted upon by gravity flexes at the knee or both extensors and flexors are severed, and attached to writing levers (Reciprocal innervation of the eye muscles is described in chapter 76)

The inhibitory effect upon the skeletal muscles is not brought about through specific inhibitory nerves as in the case of cardiac inhibition (vagus)

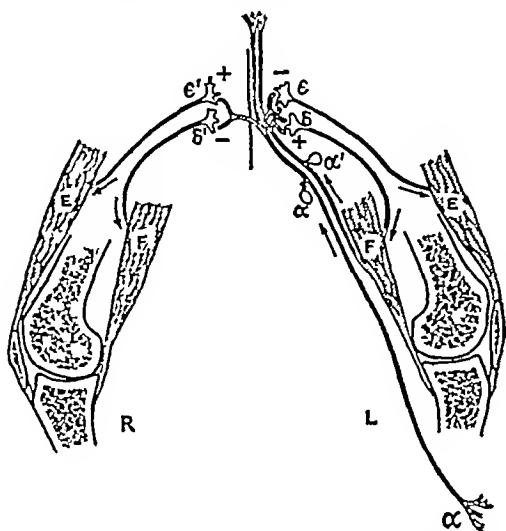


FIG 64.8 Diagram illustrating reciprocal innervation. The afferent fibers (α) from skin of leg and (α') from the flexor muscles of the knee (in hamstring nerve) pass to the spinal cord, where each divides into two. One division goes to a motoneuron (e) of the extensor muscles (E) and the other to a motoneuron (δ) of the flexor muscles (F). A branch of each afferent also crosses the mid line to similar motoneurons (e' and δ') on the opposite side of the cord. The afferent impulses, as indicated by the plus and minus signs either excite or inhibit the motoneurons, the respective effect being a function of the synapse. For the sake of simplicity, internuncial neurons between the afferent and efferent neurons have been omitted (after Sherrington)

or of the inhibition of the intestine (sympathetic). Stimulation of a muscular nerve results only in excitation. The inhibitory process is therefore central in origin. Gasser has proposed a theory for this type of central inhibition founded upon the subnormal phase of excitation in the nerve fiber. This is illustrated in figure 64.9. In order for the transmission of a nervous effect from neuron to neuron a certain minimal number of synapses must be active simultaneously, and excitation is proportional to the number. But for the sake of simplicity, let the minimal number be two. If I and II are afferent neurons, Fiber II makes two contacts with internuncial neuron a and three with b . Fiber I has only two synaptic connections with b , and two

with c . Impulses arriving by I alone, and impinging synchronously on b and c set up a discharge which is transmitted to the flexor motoneuron F. Now, when fiber II is excited a train of impulses will arrive synchronously at a and b , and since this fiber has three synapses with b its effect will override the effect of the impulses arriving over I. Let it be supposed that the impulses in the two afferent fibers have different rhythms, and that those arriving by fiber I reach the internuncial neurons in the subnormal phase caused by the transmission of the impulses over fiber II. Impulses over fiber I will, therefore, be ineffective and excitation of F will cease. But impulses arriving by fiber II will be

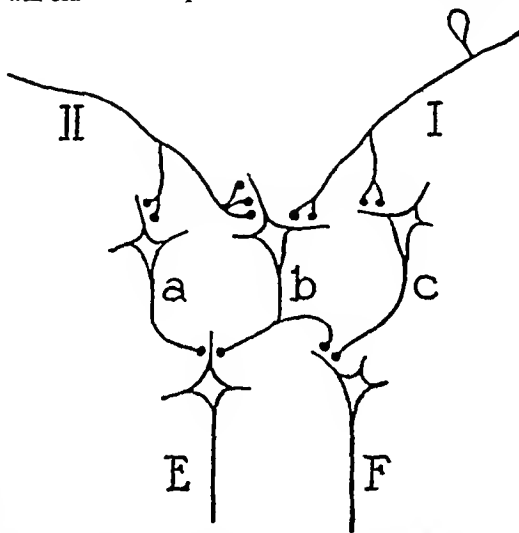


FIG 64.9 Illustrating the possible mechanism in reciprocal inhibition. (From Gasser) Description in text.

transmitted to the extensor motoneuron E. Thus, contraction of the extensor muscle and relaxation (inhibition) of the flexor muscle result.

DIRECT CENTRAL INHIBITION

The type of central inhibition described in the last section has been termed *indirect* inhibition by Lloyd. Another form termed *direct* inhibition brought about by stimulation of an afferent nerve has been discovered, it has been demonstrated to act upon reflexes involving an arc of only two neurons, and occurs when the inhibitory and excitatory impulses (in separate afferent nerves) converge upon the motoneurons. Direct inhibition cannot, therefore, be explained upon the basis of the subnormal phase in internuncial neurons. Inhibition of a two-neuron reflex lasting for over 100 m. sec. can be induced by stimulating a suitable afferent nerve. Several examples of direct inhibition have been described within recent years. The

first was observed by Lloyd, namely, the inhibition of an afferent volley in the 6th lumbar posterior root of the two-neuron reflex evoked by a synchronous afferent volley in the 1st sacral posterior root

The cause of direct inhibition is unknown, but several hypotheses have been proposed to account for it. Like indirect inhibition it is of central origin since the peripheral nerves do not contain inhibitory fibers. Nor are there any differences, so far as is known, between the central endings (synaptic knobs) which could serve excitator and inhibitory functions respectively. The most accept-

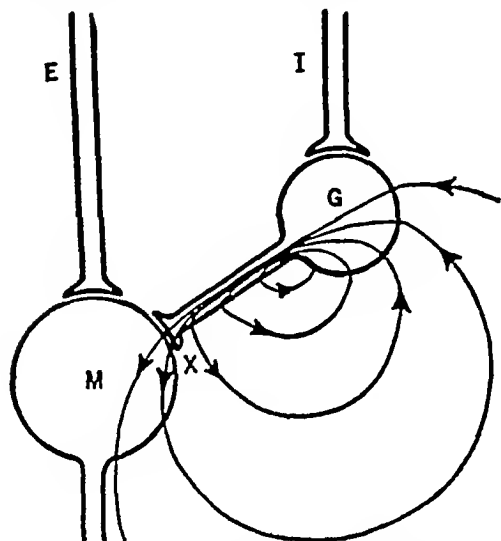


FIG 64 10 Central inhibition I, inhibitory collateral from another dorsal root ganglion cell, G, Golgi internuncial which acts to suppress by anelectrotonus ventral horn cell, M, X, anelectrotonic potential at soma of ventral horn cell (Brooks and Eccles)

able theory is one which has polarization as its basis, namely, the setting up of anelectrotonic currents in adjacent neurons activated by the inhibitory volley. The spread of the anelectrotonic effect to the motoneurons could produce a block in conduction and a cessation in the flow of impulses. A variant of this theory is one proposed by Brooks and Eccles which postulates that the short axon cells, known as Golgi Type II cells, when stimulated, though not themselves producing or transmitting impulses can generate anelectrotonic currents, and thus cause a block in conduction (fig 64 10). A chemical theory has also been proposed, viz, the liberation of an inhibitory or excitatory substance at different synaptic knobs.

Central Inhibitory and Excitatory States (*c i s* and *c e s*)

Inhibition, like excitation, undergoes summation and can be graded in intensity gradually increas-

ing the strength of an inhibitory stimulus causes a corresponding increase in the inhibitory effect upon a reflex. The principles of convergence and subliminal fringe can also be demonstrated for the inhibitory process. A central inhibitory state (*c i s*) analogous to a central excitatory state is, therefore, spoken of. In reflex action one state can be pitted against the other in the spinal center. If the *c e s* is antagonized by the *c i s* the central excitatory state is reduced or completely abolished. The *c i s* takes a certain length of time for its development to threshold value, the inhibitory state a longer time. Each persists for an appreciable period during which its intensity gradually diminishes, the *c i s* lasts for a longer time than does the *c e s*.

The *c e s* accumulated as a result of afferent impulses is dissipated by an antidromic volley, this is shown by the following experiment. When two single shocks of equal intensity and separated by a short interval are applied to an afferent nerve, the second response is greater than it would have been had it not been preceded by the first (facilitation). If between the two stimuli (and about 18 msec after the first) the motor nerve⁶ of the muscle be stimulated so as to "back fire" a volley of impulses into the reflex center, i.e., antidromically, the second response shows little or no facilitation. The annulment of the facilitation is ascribed to the inactivation of the *c e s* formed by the first afferent volley in the subliminal fringe of the motor neurons⁷. An antidromic volley has no effect upon *c i s*. The central inhibitory state is therefore looked upon simply as the inactivation of the central excitatory state, having no direct effect upon the motoneurons. In other words, if a central inhibition exists, the antidromic volley finds the *c e s* already inactivated and so can produce no further effect.

The overlap of excitatory fields has been considered (p 955). Overlap of an excitatory field by an inhibitory field is indicated by the following observation. A weak inhibitory stimulus pitted against a weak excitatory stimulus frequently fails to reduce the latter's effect. If, however, the excitatory stimulus is increased, then the weak inhibitory one is definitely effective. This result is best explained by the conception that the excitatory field has been enlarged by the stronger stimulus and so has been overlapped by the inhibitory field.

⁶ The posterior spinal roots of the nerve are sectioned in order to block any afferent impulses.

⁷ The antidromic volley, it is believed, does not pass beyond the surface film interposed between the motoneuron and the afferent terminals, i.e., the synapse. The latter is therefore considered to be the site where the accumulation of *c e s* occurs.

THE PHYSIOLOGICAL MECHANISMS GOVERNING POSTURE AND EQUILIBRIUM

STRETCH OR MYOTATIC REFLEXES

The reflex contraction of a healthy muscle which results from a pull upon its tendon is called the stretch or myotatic reflex. Any attempt to flex the limb of a decerebrate preparation, for example, meets with considerable resistance. This is due to the reflex contraction of the extensor muscles whose receptors (muscle spindles) are sensitive to a stretch stimulus (p 942). After the section of either its motor or afferent nerve the response is lost, the muscle then offering little resistance to passive flexion. In figure 65 1 are shown records of the reflex response to stretching the knee extensor

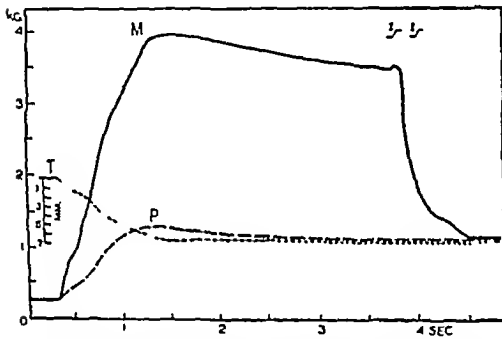


FIG. 65 1 The stretch reflex of the quadriceps (cat). Muscular response (M) before and (P) after cutting nerve to the muscle. T is a record of the table fall which stretches the muscle. At the right is shown the effect of reflex inhibition, evoked between t and t' by stimulation of afferent fibers in the ipsilateral peroneo-popliteal nerve (from Liddell and Sherrington)

(quadriceps) before and after section of its motor nerve. Liddell and Sherrington found that stretching the muscle by as little as 0.8 per cent of its original length was sufficient to evoke a reflex response. The tension developed may amount to as much as 2,000 grams. Slow as well as rapid stretching of the muscle elicits the reflex. The strength of the contraction continues to increase within limits so long as the stretching force increases, this fact, in view of the results of Adrian and of Matthews (p 944), is probably due to a rise in frequency of the discharge from individual muscle receptors (muscle spindles) as well as to more of these being stimulated—*peripheral recruitment*—the force of

the muscular contraction is thus automatically adjusted to the degree of stretching force applied.

A stretch of constant degree causes a maintained steady contraction, the muscle spindles and the stretch receptors in tendons showing very slow adaptation (p 945). The reflex ceases immediately upon withdrawal of the stretching force, i.e., it shows little or no after discharge (p 952). Its latent period is short, less than 20 msec. It is readily inhibited by stimulation of an ipsilateral or contralateral afferent nerve or by stretching the antagonistic flexor muscle, overextension in the intact animal is thus automatically prevented. The stretch reflex is obtained predominantly from those muscles which maintain the body's posture, e.g., quadriceps, soleus, gastrocnemius and other extensors. In the decerebrate animal flexors show little or no response to steady stretch, but in the intact or thalamic animal the flexors also exhibit these reflexes (see supporting reactions). The labyrinthine and neck reflexes exert their effects upon the limb muscles, chiefly, through the reinforcement of myotatic reflexes.

The afferent fibers of the myotatic reflex arcs are large calibered, myelinated, rapidly conducting fibers of the A group. The dark muscle fibers of the extensors, upon which the maintenance of posture depends, are more plentifully supplied with these large diameter nerve fibers than are the pale fibers of flexor muscles. The stretch reflexes show reciprocal innervation, when a muscle reacts by contraction to a stretching force its antagonist relaxes. The *myotatic unit* which consists of, tendon or muscle receptor, afferent nerve fiber, spinal center and motoneuron, "exhibits within itself in full measure the elementary mechanism of reciprocal innervation" (Lloyd).

THE KNEE AND ANKLE JERKS

A sharp tap upon the patellar tendon, when the knee joint is semi-flexed and the leg dependent, causes a short twitch of the knee extensor. Also, a blow upon the tendo Achillis causes a quick contraction of the calf muscles. These brief contractions or jerks, as well as those described below for the upper limb, are "fractional examples of

the stretch reflex" and are not due to stimulation of the tendon or to direct stimulation of the muscle. The tap causes a sudden momentary stretch of the muscle (a stretch of 0.05 mm or less in $\frac{1}{10}$ second being sufficient to elicit the response). Rendering the tendon itself insensitive by means of cocaine does not prevent the reflex. A voluntary action, such as clapping the hands together, reinforces the reflex and increases the force of the jerk. *Patellar* or *ankle clonus* consists of a rapid rhythmical series of twitches of the quadriceps or calf muscles, respectively, which results from the application of a sudden stretch stimulus when, as in a lesion of the pyramidal tract, the reflex arc concerned is abnormally excitable. A sharp depression of the patella by the examiner or a sharp dorsiflexion of the ankle elicits the respective clonic response. The center for the knee jerk in the human subject lies in the 2nd, 3rd and 4th lumbar segments of the cord, it is innervated through the anterior crural nerve. The center for the ankle jerk is situated in the 1st and 2nd sacral segments, the peripheral nerve is the sciatic. Lloyd has shown that only two neurons are involved in the reflex arc of the knee and ankle jerks, and of other myotatic reflexes, i.e., the impulse passes from afferent fiber to motor neuron without traversing an internuncial neuron.

For a long time the knee jerk was thought to be a direct response of the muscle to mechanical stimulation and not a reflex, since its latent period was only some 6 to 12 msec. The revised estimates of the speed of conduction in peripheral nerve (up to 120 meters per second) show that this time is not too short for the occurrence of a purely spinal reflex, which the knee jerk is now generally accepted to be. Any condition, such as decerebration or injury to the pyramidal tracts (ch. 66), which enhances the stretch reflexes increases the tendon jerks.¹ They are abolished by an injury or disease involving the efferent or afferent limb of the reflex arc or the center itself (anterior horn cells). On account of its brief twitch-like character the knee jerk is sometimes spoken of as the "phasic reaction" of the stretch reflex, whereas the sustained

contraction resulting from a continuous pull upon the tendon, and which is concerned in the maintenance of posture, is referred to as the "static or postural reaction" of the stretch reflex. The knee jerk is less subject to abnormal states, e.g., spinal transection, anesthetics, circulatory failures, etc., than are the postural reactions. In an animal such as the dog or rabbit, the knee jerk returns in a few minutes after spinal transection, in the monkey after some days. In man, though postural reflex activity is entirely lost, the muscles being quite flaccid, the knee jerk is elicitable in some degree in from 3 to 7 weeks after a complete transverse lesion of the cord (see ch. 66).

The smooth and steady character of the contractions brought about through the ordinary stretch reflexes concerned with the maintenance of posture, is due to the asynchronous nature of the impulses set up by the numerous stretch receptors and the asynchronous reflex discharge of impulses down the motoneurons to the muscles. The muscle fibers are, therefore, never all relaxed nor all contracted at the same time. The sharp contraction of the quadriceps characteristic of the knee jerk, on the contrary, is brought about by the discharge of a volley of synchronous impulses and therefore by the fibers of the muscle contracting in unison.

Ankle clonus. Under certain conditions a muscle or group of muscles instead of contracting smoothly and continuously may do so intermittently in a series of rapidly repeated movements or jerks. Thus, in lesions of the central nervous system associated with hypertonus, sharp passive dorsiflexion at the ankle joint with maintenance of the foot in the dorsiflexed position by light pressure upon the sole, evokes a stretch reflex which consists of a series of rhythmical contractions of the stretched muscles. The sudden dorsiflexion causes a sufficient number of stretch receptors to discharge a synchronous volley of impulses along the afferent nerve to the spinal centers and a synchronous volley to be discharged to the muscles. The continued stretch of the muscle caused by keeping the ankle in the dorsiflexed position sets up another synchronous volley as soon as the motoneurons have recovered each time from the subnormal phase of the preceding one. Thus, the clonic movements continue so long as the muscles are kept upon the stretch. Clonus can be evoked in a similar fashion in other situations.

¹ When the electrical responses are recorded from the muscle of an animal during the elicitation of a tendon jerk, cessation of the action currents is found to occur during the actual contraction of the muscle. Thus, the so called "silent period" is probably due to the muscle spindles being relieved of stretch as the muscle shortens (see p. 944). An additional factor is the synchronous nature of the discharge down the motoneurons, for, the latter being all excited in unison pass also simultaneously into the subnormal phase.

SOME OTHER STRETCH REFLEXES (DEEP OR TENDON REFLEXES) OF CLINICAL IMPORTANCE

The *jaw jerk* Tapping the chin with the mouth partly open, and jaw supported, stretches the masseters which contract and jerk the jaw. The center is in the pons.

The *biceps jerk* is elicited by a sharp tap upon the biceps tendon in front of the elbow joint, the response consists of a quick contraction of the biceps with flexion of the elbow. The center for the reflex is situated in the 5th and 6th cervical segments of the cord, it is innervated through the musculocutaneous nerve.

The *triceps jerk* is evoked by a blow upon the triceps tendon just above the olecranon process of the ulna, contraction of the muscle and extension of the elbow result. The center for the response lies in the 6th and 7th cervical segments, the peripheral nerve is the musculospiral (radial).

The *supinator jerk* consists of contraction of the supinator muscle and flexion of the elbow, it follows a blow upon the styloid process of the radius. The center lies in the 5th and 6th cervical segments of the cord, the peripheral nerve is the musculospiral.

Rossolimo's reflex which is seen in lesions of the premotor cortex or of extrapyramidal fibers, consists of flexion of the toes, including the hallux, when the toes are flicked. The same response is elicited by tapping the dorsal surface of the foot over a metatarsophalangeal joint and is then known as the *sign of Mendel-Bechterew*. A response corresponding to that of Rossolimo can be evoked by flicking the finger tips.

These are all myotatic reflexes caused by the sharp stretching of muscles, e.g., the flexors of the fingers or toes, biceps, triceps or closures of the jaw.

THE TONE OF SKELETAL MUSCLE

As a result chiefly of the work of the Sherrington school the word "tone" or "tonus" as applied to skeletal muscle has acquired a clearly defined meaning which it had previously lacked. Muscle tone is the steady reflex contraction which resides in the muscles concerned in maintaining the posture characteristic of a given animal species. To use Sherrington's words "reflex tonus is postural contraction." Tonus has its basis in the "static reactions" of the stretch reflexes, and its seat is therefore mainly in the *antigravity muscles*. In most mammals these are extensor muscles and we shall see that in decerebrate rigidity the animal exhibits an attitude which is a caricature of

standing, due to an exaggeration of the tone of the extensors.

In man the antigravity muscles and consequently those which exhibit the greatest degree of tone are the retractors of the neck, the elevators of the jaw (masseters), the supraspinatus, the extensors of the back, the ventral muscles of the abdominal wall (probably), and the extensors of the knee and ankle (vastocruureus, gastrocnemius and soleus). When these muscles are completely relaxed, as in an unconscious person, the body collapses. In the healthy conscious person, stretch reflexes are largely instrumental in preventing this occurrence for, as we have seen, the elongation of an antigravity muscle by less than 1 per cent of its length stimulates the muscle spindles and a sustained reflex contraction results (see also supporting reactions, p. 970).

Though the fundamental basis of tonus in voluntary muscle is the myotatic reflex centered in the cord, the tonus state is influenced profoundly by higher centers. Impulses from labyrinthine and neck muscle receptors (p. 969) exert their influence upon this background of tonus established through lower spinal centers. Similarly, pathways from cerebellar, midbrain and cerebral centers convey impulses which, impinging upon the final common path, are capable of altering the degree of tonus, of effecting finer adjustments in the tonus state and of maintaining its normal distribution between groups of muscles (fig. 646, p. 957). The tone of a given group of muscles may also be influenced through the spinal centers by impulses arising in other muscle groups (e.g., neck muscles and the muscles of the digits, ankle and wrist, as in the positive supporting reaction) and in skin receptors.

When the spinal centers are separated from higher centers, extensor tone is abolished (see spinal shock, p. 967). Also, since it is dependent upon stretch reflexes, muscular tone is lost after injury to the afferent or efferent nerves, or to the spinal centers.

A feature of tonic contraction is its economy in the expenditure of energy. Posture is maintained for long periods with little or no evidence of fatigue, e.g., in decerebrate rigidity, in the maintained closure of the jaws, standing or sitting in the normal person, and in the clasping reflex of the frog. The increase in metabolism is less than in the case of those ordinary contractions which result in movement, though the difference is not as great as was once supposed.² It had also been

² In the normal human subject postural contractions cause a rise of from 50 to 70 per cent in the basal metab-

thought that tonic contractions were unaccompanied by action currents. This, however, has been shown to be an error. The work of Forbes and of Adrian and their associates indicates that the economy of energy is effected through different groups of muscle fibers contracting in relays, only a proportion of the total number of fiber groups of the muscle being active at any moment. Thus, active fiber groups mingled with inactive groups are scattered throughout the muscle. The alternating periods of rest and activity of the muscle groups explains the ability of the tonic contraction to be maintained for so long without showing fatigue.

RED AND WHITE MUSCLE

The skeletal muscles of many animals, e.g., birds, rabbit, cat, etc., can be clearly divided into two types, (a) *red or dark*, and (b) *white or pale*. The former, as compared with the latter, are composed of smaller fibers having a granular and more opaque appearance, possessing more distinct longitudinal striations but less pronounced cross striations, and containing a larger proportion of sarcoplasm.³ The *red* fibers contract more slowly, fatigue less readily, and are tetanized at a slower rate of stimulation than are the pale fibers. The *pale* fibers are translucent, show prominent cross striations and a small quantity of sarcoplasm. Those muscles which execute rapid movements are usually, though not invariably of the pale variety, whereas the slower movements are carried out chiefly by the dark muscles. It is probable that all muscles are a mixture of the two types of fiber, but that in some the red, in others the pale type, predominates. The segregation of the two types of fiber in different muscles is much more pronounced in some animals, e.g., the fowl and rabbit, than in others. In birds capable of soaring, or of long flight the wing muscles are mainly of the red type. In man and the monkey the two kinds of muscle can be distinguished with the naked eye, though the differences are not very pronounced. The predominance of one or other type of fiber, can, however, be readily made out under the microscope. The rapidly contracting flexor muscles are largely composed of the pale fibers, whereas the extensors have usually each a superficially placed, rapidly contracting, pale component (or head) and a deep, slowly acting dark head. For example, the pale

gastrocnemius contracts rapidly and overlies the dark slowly contracting soleus. The lateral short head of the triceps is pale and contracts rapidly, the medial short head is dark and contracts more slowly. The slowly acting deep components of the extensor muscles are those which respond most readily to a stretch stimulus by a prolonged steady contraction and so show the greatest degree of tone. They are the "focus of the stretch reflex," and it is upon them that labyrinthine and neck reflexes are mainly exhibited. The rapidly contracting portions are more especially concerned with the execution of rapid movements. No hard and fast line, however, can be drawn, since both types of extensor muscle exhibit tone and both take part in phasic contractions, and the pale muscles respond to stretch but by a sharp contraction as in the tendon jerks, and in the "pluck reflex" of a flexor muscle, which consists of a brief contraction when a sharp jerk is applied to a flexor or tendon.

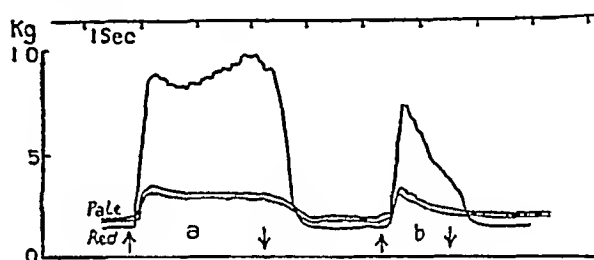


FIG. 65.2 The reflex effect of the labyrinthine and neck reflexes on red and pale muscle. Preparation with section of brain-stem slightly anterior to superior colliculi: M. triceps, short lateral head (pale), double traced line; short medial head (red), single line. Neck dorsiflexed at ↑ and ventriflexed at ↓. Labyrinth in intermediate position in each neck posture (from Creed, Denny-Brown, Eccles, Liddell and Sherrington *Reflex Activity of the Spinal Cord*, Clarendon Press)

THE QUESTION OF THE SYMPATHETIC INNERVATION OF SKELETAL MUSCLE

The apparent dissimilarities between postural or tonic contraction and the rapid contractions resulting in movement (phasic contractions) have led some physiologists to believe that these two types of muscular activity are fundamentally different. A theory was built up which postulated two components in muscular activity. A *plastic* element (plastic tonus) sustained by the sarcoplasm of the muscle fiber, and dependent upon a sympathetic innervation and a *contractile* element (so-called contractile tonus) involving the myofibrils and governed by somatic nerves. The ordinary phasic contractions, it was supposed, were dependent solely upon the myofibrils and somatic innervation, the sustained or tonic contraction upon the sarcoplasm and sympathetic innervation. The red muscles with a large proportion of fibers rich in sarcoplasmic material were thought to be innervated by many sympathetic fibers. Sympathetic denervation of spastic muscles was,

olism. In certain nervous states, in which the body is maintained in fixed attitudes for long periods (catatonias), the increase in metabolism resulting from the tonic contraction of the muscles is less (20 per cent) than in the case of the normal subject. The circulatory effects of sustained posture are also less pronounced in the pathological cases (Gaylor and Wishart, *Braun*, 1933, 56, 282).

³ They resemble the intrafusal fibers of the muscle spindle (p. 942).

therefore, resorted to in attempts to relieve the spasticity. The Orbeli effect also appeared to support the belief in a sympathetic innervation of skeletal muscle. Orbeli found that the fatigue of a muscle stimulated *through its nerve* (i.e., indirectly) was reduced by concurrent stimulation of sympathetic fibers going to the muscle. Sympathetic stimulation was, however, without effect upon the contractions of a muscle stimulated *directly*. These results lead to the conclusion that the sympathetic effect was exerted upon the motor end plate. Corkill and Tiegs confirmed Orbeli's observations. Nevertheless, no sympathetic fibers have been demonstrated as terminating in the motor end plate,

between the superior colliculi and the vestibular nuclei.⁴ It was first studied and described in detail by Sherrington. The animal assumes a characteristic attitude with limbs stiffly extended, head retracted, jaws closed, and tail horizontal or erect. When placed upon its feet the limbs support the weight of the body (fig. 65.4A). *The position is a caricature of the normal standing position.* The knee jerk and other stretch reflexes (p. 960) are exaggerated. The righting reflexes are abolished, tonic neck and labyrinthine reflexes are retained. The ability to regulate the body temperature is lost

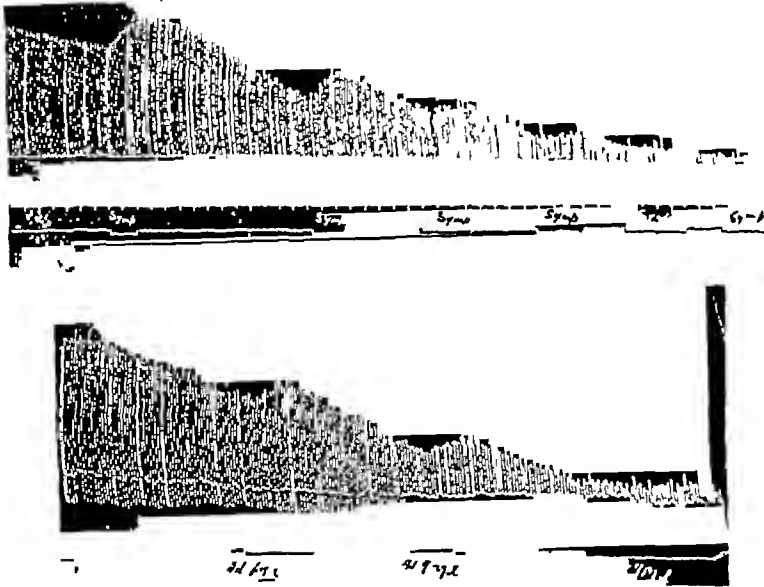


FIG. 65.3 The effect of sympathetic stimulation (upper tracing) and of adrenaline (lower tracing) upon the fatigue of skeletal muscle activated through its motor nerve (from Corkill and Tiegs)

the effect, it appears, results from the liberation of an adrenaline-like substance (sympathin) from vascular sympathetic endings (fig. 65.3). The action of adrenaline in postponing muscular fatigue has been mentioned, ch. 59.

The conception of a sympathetic innervation of skeletal muscle has been discredited as a result of these findings, and the complete lack of evidence which can be brought to its support. Excision of the sympathetic nervous system does not alter any postural contraction.

DECEREBRATE RIGIDITY

This is the term applied to the sustained contraction of the extensor muscles which supervenes upon transection of the brain stem at any level

In some instances there is increased tone of the flexors as well, but the characteristic feature of decerebrate rigidity is the tonic contraction of the muscles which maintain the posture of the body against gravity (p. 963), the *antigravity muscles*. It may be pointed out that in the frog whose natural posture is squatting with flexed thighs, legs and arms, it is the flexor muscles which are the site of decerebrate rigidity. Also in the sloth, whose habit it is to remain for long periods suspended from a tree branch, flexor rigidity is the

⁴ Decerebration can also be produced by tying the common carotids and the basilar artery at the center of the pons and thereby depriving the fore-brain of its blood supply.

characteristic result of decerebration. In the pigeon the flexor muscles which maintain the resting (folded) position of the wing exhibit rigidity, and in the ape the muscles which hold the elbow semiflexed when the body is erect show increased tone. A remarkable feature of decerebrate rigidity is its maintenance for many hours without the muscles showing signs of fatigue.

TONIC LABYRINTHINE AND NECK REFLEXES IN THE DECEREBRATE PREPARATION (SEE ALSO P 969)

Labyrinthine proprioceptors are responsible for the tone of the neck extensors, consequently, after destruction of the labyrinths in the decerebrate preparation the head is no longer held erect but falls into the fully flexed position. The rigidity in

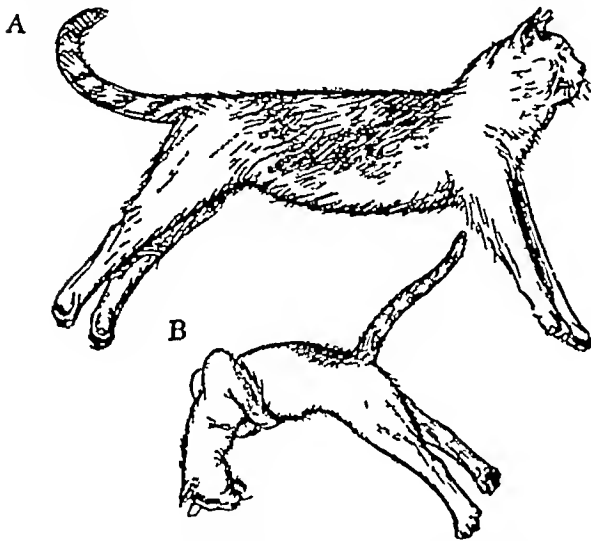


FIG 65.4 A suspended decerebrate cat showing extensor rigidity. In A, the labyrinths are intact. B, a decerebrate and labyrinthectomized animal. As a result of the destruction of the labyrinths the head has dropped, this position of the head induces reflex flexion of the forelimbs and extension of the hindlimbs (tonic neck reflexes, p 971) (after Pollock and Davis)

the forelimbs in turn is maintained through proprioceptive reflexes initiated in the extended neck muscles as well as directly through labyrinthine reflexes (p 969). After labyrinthectomy, therefore, the flexed position assumed by the head sets up proprioceptive impulses from the neck muscles which lead to a reduction in the extensor tone of the forelimbs, these then become strongly flexed upon the chest (fig 65.4B). If the neck muscles of the labyrinthectomized animal are deafferented, movements of the head are without effect upon the extensor muscles of either fore or hind limbs, deafferentation of the forelimb muscles themselves is followed by little change in the extensor tone of the forelimbs. On the con-

trary, the extensor rigidity of the hind limbs is affected relatively little by labyrinthectomy or deafferentation of the neck muscles, but is abolished by deafferentation of the hind limb muscles themselves. To sum up, the tone of the forelimb extensors is maintained chiefly by labyrinthine and neck reflexes, proprioceptive impulses from the forelimb muscles themselves playing a very minor part. In the hind limbs the extensor rigidity is sustained *mainly* through proprioceptive reflexes initiated from the corresponding limb muscles.

LENGTHENING AND SHORTENING REACTIONS

Attempts at passive flexion of the extended limb of the decerebrate (or chronic spinal) preparation are met by considerable resistance. If the force is maintained inhibition of the stretch reflex (upon which the resistance depends) results, the limb gives way suddenly (*clasp-knife effect*) and may then be easily flexed to any degree and remains in the new position. The elongation of the extensors which permits the flexion of the limb is called the "*lengthening reaction*". Upon moving the limb again into the extended position the extensor muscle shortens adaptively, and again resists the limb being bent. This is the "*shortening reaction*". These reactions give the muscles a certain plastic quality. The giving way of the extensor spasm in the lengthening reaction has been ascribed to the setting up of inhibitory proprioceptive impulses in the muscle when the stretch stimulus reaches a certain intensity. As the knee extensor is inhibited by a contraction of the opposite knee extensor occurs, this is known as *Phillipson's reflex*.

A CONSIDERATION OF THE CENTRAL MECHANISMS CONCERNED IN THE PRODUCTION OF DECEREBRATE RIGIDITY

Bazett and Penfield succeeded in keeping animals (cats) alive for from two to three weeks after aseptic section through the mid-brain. The animals showed extensor rigidity up to the day of death, though as time elapsed the condition tended to be less constant, intervals of reduced extensor tone and increased flexor tone alternating with periods showing the typical decerebrate posture. The fact that extensor rigidity was present so long after operation makes it clear that the condition is not due to the irritation of fibers at the line of section, for these must have undergone degeneration before the end of the survival pe-

mod. Decerebrate rigidity is evidently due to the release from higher control of a center or centers situated below the level of the transection. The rigidity is increased after ablation of the anterior lobe of the cerebellum, the stretch reflexes of the antigravity muscles are exaggerated. On the other hand, stimulation of the cortex of this part of the cerebellum causes pronounced inhibition of decerebrate rigidity. The rigidity is abolished by destruction of the vestibular nuclei or of the vestibulospinal tracts. Intact vestibular nuclei, therefore, are essential for the development and persistence of decerebrate rigidity. Section confined to the pyramidal tracts has no important effect upon decerebrate rigidity, but unilateral section of the medulla above the decussation of the pyramids abolishes muscular rigidity on the *same* side, whether or not the pyramidal fibers are excluded from the section.

Some of the pseudoaffective reactions characteristic of the thalamic or decorticate animal (p 967) are seen in the decerebrate preparation to a minor degree. Bazett and Penfield record that in their chronic decerebrate animals stimulation of the preparation, as by twisting the ear, caused lashing of the tail, licking and running movements were induced by merely touching the animal. Purring followed the introduction of milk into the stomach. Other reflexes whose centers are apparently located in the medulla consisted of motions of chewing and swallowing upon the passage of a stomach tube, and licking movements upon its removal, growling, mewling or crying also occurred in response to various types of stimulus. Auditory reflexes, consisting of movements of the head, limbs or tail, in response to certain sounds, especially to a small scratching sound resembling that made by a mouse were also observed.

A decorticate cat or dog (thalamus, hypothalamus and basal ganglia intact) shows a nearly normal distribution of muscular tone, but the decorticate monkey develops well marked hypertonus, though not typical decerebrate rigidity. The hypertonus is increased if the section excludes the hypothalamus and thalamus. Only a moderate hypertonus is produced by destruction of the red nuclei.

Excision of certain areas of the cortex—so-called suppressor areas (ch. 68)—increases muscle tone. A suppressor area is also situated in the medulla which is subordinate to the cortical areas, and which upon stimulation inhibits decerebrate rigidity. On the other hand, a long tract is situated in

the brain stem (facilitatory tract) whose destruction reduces spasticity though does not abolish it. It is evident then that decerebrate rigidity and muscle tone in general cannot be simply explained, nor be accounted for upon the basis of known function of a limited part of the central nervous system. The most important nervous mechanisms involved in the production of decerebrate rigidity appear to be the release of the vestibular nuclei, and the facilitatory tracts from suppressor influence.

DECEREBRATE RIGIDITY IN MAN

States of hypertonus resembling decerebrate rigidity are seen in certain nervous diseases. The spasticity of the hemiplegic limbs, for example, has been compared with the experimental condition, but the hypertonus is, as a rule, much less pronounced in the hemiplegic patient than in the decerebrate animal. Moreover, in hemiplegia, though the lower limb is extended, in the arm it is the flexor muscles which are hypertonic. It may be mentioned, however, in this connection that in the ape standing upon its hind limbs the flexors of the forelimbs are the postural muscles and hold the limbs in a semiflexed position against gravity, it is the antigravity muscles, as already mentioned, which show rigidity after decerebration. When, on the other hand, the animal goes upon all fours the extensor as well as the flexor muscles of the forelimb become hypertonic like those of lower mammals and convert the limbs into rigid supports for the body. It has been shown by Brahm that when the hemiplegic patient assumes the quadrupedal position the rigidity also shifts from the flexors of the arm to the extensors, the arm is extended and offers resistance to passive flexion. This suggests that the ordinary flexed position of the hemiplegic arm is in reality, as in the decerebrate ape, a part of reflex bipedal standing.

In extensive destructive lesions of the cortex or subcortical connections of the frontal lobes and not confined to the motor area (see p 967), manifestations similar to those seen in the thalamic (decorticate) monkey appear. Tonic labyrinthine and neck reflexes can be readily demonstrated (fig 65 6, p 968).

Decerebrate rigidity in man more closely comparable to that seen in animals is seen as a temporary or functional manifestation in certain convulsive states—tonic fits—and as a permanent state in lesions of the pons or mid-brain. All four limbs are rigidly extended, with the arms thrust

strongly backwards and the forearms pronated. The muscles show the lengthening and shortening reactions characteristic of the decerebrate animal and tonic labyrinthine and neck reflexes can be demonstrated in a striking manner.

Rigidity is a pronounced feature of striatal disease but it differs from decerebrate rigidity in that the flexors as well as the extensors are hypertonic.

THE SPINAL STATE

Transection of the cord produces an immediate flaccid paralysis of the muscles behind the point of section. Immediately after section in the lower cervical region the limbs hang limply, the extensors being quite toneless, the stretch reflexes and other extensor responses cannot be elicited, the knee jerk is abolished. The blood pressure falls to a dangerously low level and vascular and visceral reflexes are unobtainable. This condition is called *spinal shock*. Its duration varies with the species. The higher the position of the latter in the phylogenetic scale the more profound is the shock and the slower is the recovery. In the frog its duration is brief, in the rabbit, cat and dog the knee jerk returns within a few minutes (rabbit), or in an hour (cat) or longer (dog), but in the monkey, not for several days. Other extensor reflexes, however, remain in abeyance for a much longer time. In the cat and dog the picture immediately following spinal transection is the converse of that seen after decerebration. The flexor reflexes though depressed for a short period soon recover or show little or no impairment, indeed, flexor responses may be actually augmented. Consequently, if the cord of a decerebrate preparation is sectioned, the exaggerated extensor tone characteristic of the latter is replaced (behind the section) by an imbalance in favor of the flexor muscles.

Spinal shock also follows section through the medulla below the vestibular nuclei (*decapitate preparation*).

In the cat and dog spinal shock is gradually recovered from over a period of weeks. The blood pressure is restored to normal and the vascular reflexes can again be obtained. The reactions of the extensor muscles return, and the animal is able, though imperfectly and for a few minutes only, to support the weight of the body when placed upon its feet (chronic spinal animal).

Spinal shock is attributed to the removal of impulses which in the intact animal descend from higher centers to reinforce the spinal centers. That it is due to this and not simply to an inhibitory effect of the local

injury itself seems clear from the fact that after an animal has recovered from spinal shock, a second transection made behind (lower than the original one) does not cause a return of the shock state. In the cat and dog the flexor reflexes are evidently dependent only to a minor extent upon the higher centers, since they are capable of being executed by the spinal centers alone. The loss of the extensor reactions results from the removal of vestibular impulses (severance of vestibulospinal tracts) but these reactions can also eventually be carried out by the spinal centers. In *primates*, on the contrary, in which a greater degree of motor function is represented in the cerebral cortex, the immediate effect of cord section upon reflex activity is much more profound, involving the flexors as well as the extensors. The loss of the flexor reactions is probably due to the greater dependence of these higher animals upon descending (corticospinal) tracts as compared with lower orders. Section of these tracts is a more important factor in this respect than is section of the vestibulospinal tracts. Ultimate recovery of reflex activity in the monkey is slight, any which results is confined to the flexors and the knee jerk. The muscles waste and the limbs remain flaccid. The failure of recovery eventually to occur is probably due to degeneration of anterior horn cells below the level of the cord section (isolation dystrophy), the long duration of the shocked state permitting deterioration of the cells to proceed beyond the time within which repair is possible (see paraplegia in flexion, ch. 66).

THE "THALAMUS" ANIMAL

This term is given to an animal whose cerebral hemispheres have been removed, leaving the optic thalami intact (see fig 65 10, Section I). Such preparations retain their righting reflexes and can regulate their body temperature. They are also capable of carrying out coordinated reflex acts and show, often to an exaggerated degree, reactions which in the intact animal are associated with emotional states (fright, anger). Such reactions, which are termed pseudoaffective, are also exhibited by an animal whose cortex alone has been ablated (see ch 67). In contrast to the decerebrate animal, the distribution of muscular tone in the thalamic cat or dog shows little departure from the normal, though pronounced extensor rigidity becomes evident when the animal is held suspended in mid-air.

Primates, on the contrary, are unable to walk after bilateral decortication, and there are profound alterations in the muscular tone of the limbs. The animal assumes a characteristic attitude. When lying upon its side the limbs of the under part of the body are extended and the uppermost limbs flexed. The uppermost limbs show a pronounced grasp reflex. Turning the animal on to its

opposite side reverses the picture. The limbs which were underneath and extended are now flexed and show the grasp reflex, while the limbs previously uppermost, being now underneath, are extended. Righting reflexes (except those dependent upon vision) and the tonic labyrinthine and neck reflexes of Magnus and de Kleijn are shown in a striking fashion (figs 65 5 and 65 6). When a board

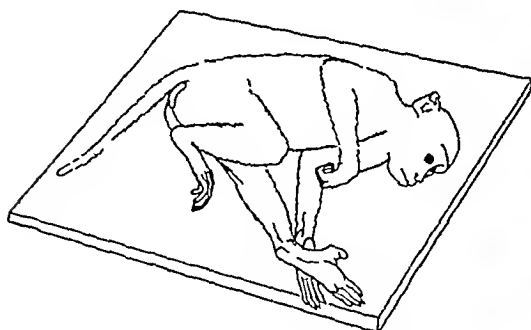


FIG 65 5 Showing posture of a "thalamus" monkey. The underneath limbs are extended and the uppermost flexed. When the animal is turned over, the previously extended limbs (being now uppermost) are flexed, and those which were flexed (now underneath) are extended (After Bieber and Fulton)

is laid upon the animal's upper surface, so as to equalize the pressure on both sides of the body, the limbs assume symmetrical positions (p. 969)

STATIC AND STATO-KINETIC REFLEXES OF MAGNUS AND DE KLEIJN

The reflex mechanisms governing the orientation of the head in space, the position of the head in relation to the trunk and the appropriate adjustments of limbs and eyes to the position of the head, are called into action by afferent impulses discharged from receptors situated in (a) the vestibular apparatus (semicircular canals or utricle), (b) the neck muscles, (c) the retina, and (d) in the body wall or limb muscles.

The postural reflexes are classed into two main groups, *static* and *kinetic*.

(1) *Static reflexes*. These are *general* and *local* or *segmental*. The general static reactions, so called by Magnus because they involve the entire body, or at least four limbs, are (a) the *righting reactions* and (b) the *stato-tonic* reflexes. The general static reactions are called into play by the deviation of the head from its usual or "natural" position in space (stimulation of receptors or utricle, p. 983) or in relation to the trunk (stimulation of neck

muscle receptors). The local or segmental static reactions involve essentially one limb or a pair of limbs, examples are, the *supporting reactions* and the *crossed extension reflex*. The latter reflex has been described on page 957.

(2) *Kinetic (stato kinetic or accelerator) reflexes*. The actual movements of the head bring these reflexes into action (stimulation of receptors in semicircular canals or utricle).



FIG 65 6 Decorticate child. Suprasellar cyst causing interruption of subcortical connections, complete quadriplegia with marked extensor rigidity of the lower limbs and semiflexed pronated position of the upper. Turning the head to one side causes extension of the arm of that side and increased flexion of the opposite arm (After L. E. Davis)

When an animal's head moves in space as a result of a movement of the neck, or of a change in the position of the body as a whole, the reflex adjustments of limbs and eyes which accompany a particular movement of the head are brought about through stato kinetic reactions. The attitude which is thus struck is *sustained* by a stato-tonic reflex so long as the head is in the new position. The righting reflexes serve to maintain the normal upright position of the body or to restore this position if as a result of some untoward movement or the application of an external force

the animal is thrown upon its side or back. A more detailed account of these different postural reflexes follows.

GENERAL STATIC REFLEXES

THE RIGHTING REFLEXES

The orientation of the head in space, and the ability to maintain the body in a certain definite (normal) relation to the head is a characteristic of animal life. A cat held back downwards and then allowed to fall through the air lands upon its feet, its head and body assuming, in a flash, the normal attitude. A fish resists any attempt to turn it from its natural position and if placed in the water upon its back flips almost instantly into the normal swimming position. Even a cray-fish rights itself from the back-down position.⁵ These righting reactions are complex and involve five separate types of reflex.

(a) Labyrinthine righting reflexes acting upon the neck muscles

(b) Neck righting reflexes acting upon the body

(c) Body righting reflexes acting upon the head

(d) Body righting reflexes acting upon the body

(e) Optical righting reflexes

The first four of these are demonstrated best upon a "thalamus" animal (p. 967). When a "thalamus" rabbit or a blindfold normal animal is suspended from the pelvis (fig. 65.7) the head turns until it assumes its normal position in space, i.e., into the position it would occupy were the animal in its natural position. The maintenance of the head in the new position is due to labyrinthine righting reflexes acting upon the neck muscles.

Turning the body of the animal through the air into different positions is followed by compensatory movements of the head, its orientation in space being thereby maintained. After extirpation of the labyrinths or destruction of the utricle alone, and suspension of the animal as before, the head shows no compensatory movements; it hangs limply like that of a dead rabbit.

When the blindfold or thalamic animal is laid resting upon its side on a table the head is raised into the usual upright position as a result of the

labyrinthine reflexes just mentioned. The contraction of the neck muscles which rotate the head sets up, in turn, proprioceptive impulses, which through a center in the upper cervical cord exert an influence upon the muscles of the body which rotate it (thorax first, then pelvis) into the normal relationship to the head. This is the *neck righting reflex acting upon the body*. A labyrinthectomized animal when laid upon its side behaves in a somewhat similar manner. The reaction under the latter circumstances is due, however, to the asymmetrical stimulation (pressure of one side of the body upon the table) of exteroceptors in the body wall, and the reflex contraction of the neck muscles. This is the *body righting reflex acting upon the head*. If

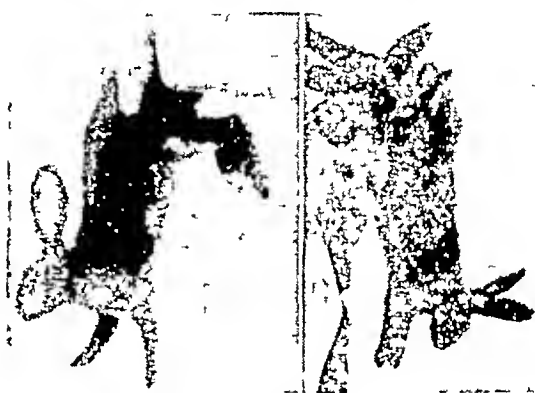


FIG. 65.7 On left, position taken up by a thalamus rabbit with intact labyrinths. As it possesses the labyrinthine righting reflexes, it carries its head in the normal position. On right, position taken up by a rabbit like the preceding but deprived of its labyrinths. The head is not raised towards the normal position (after Magnus).

a board of a weight equal to that of the animal is laid upon its upper surface, the pressure being thereby disposed equally on the two sides, the compensatory movement of the head does not occur.

Again, when a blindfold animal or one which has been decorticated is laid upon its side, and its head held down to the table, to eliminate the righting reflexes of labyrinthine and cervical origins, as well as the body righting reflex acting upon the head, the body nevertheless attempts to right itself by raising the hind quarters. This must be due to the asymmetrical pressure upon the body exerting a reflex effect upon the skeletal musculature, i.e., a *body righting reflex acting upon the body*.

This reflex is well shown by the labyrinthectomized dogfish, especially after blinding. When placed in the water (pressure being equal on all sides) upon its side or back the fish deprived of its

⁵ The otocyst of the prawn is homologous to the utricle of mammals. It is open to the exterior and at moulting time its lining is shed. The otoliths consist of grains of sand which the animal itself introduces into the otocyst. If, after moulting, the animals are placed in a dish containing fine iron filings these are inserted and various forced movements of the head and body can be induced by bringing a magnet into relation with the ear.

labyrinths swims away in the false position. When, however, it comes into contact with the bottom or side of the tank (pressure then being exerted unequally) the righting reaction immediately occurs. Even worms possess this means of orientation.

The *optical righting reflexes* are initiated through retinal impressions. They are absent in the thalamic animal since their center is cortical (occipital lobes). Visual impressions play a prominent role in the orientation of the head in some animals, such as the monkey, dog and cat. If a labyrinthectomized dog is held in the air in order to exclude the body righting reflexes, it is capable of orienting the head, but is not able to do so if blindfold. The optical righting reflexes are of minor importance as compared with the labyrinthine in animals such as the rabbit and guinea pig whose cortical development is more rudimentary, and who depend to a large extent upon subcortical visual centers. For this reason, in the more intelligent animals the optical righting reflexes are abolished by decortication. Yet even the cray-fish deprived of its otocyst, which is the homologue of the utricle of mammals, rights itself, though with less facility than normally, when placed upside down in water. In this medium righting reflexes due to unequally distributed pressures upon the body surface must obviously be in abeyance, the righting reaction, then, is apparently entirely of retinal origin. That this is so is proved by blinding the animal, then the righting reaction is lost completely.

To sum up, the righting reflexes may now be given in their natural sequence. When the animal is placed upon its back the labyrinthine reflex acting upon the neck muscles turns the head into its normal relationship to the dimensions of space, the proprioceptive reflexes of the neck muscles then bring the body into its normal relation to the head. When resting upon a rigid support these reflexes are reinforced by body righting reflexes (on head and body). When the animal falls through air or water, these latter reflexes of course do not come into play. A labyrinthectomized but otherwise normal animal, such as the cat or dog, recovers its upright position when allowed to fall through the air, as a result of the operation of the optical righting reactions, the righting ability is lost, however, if the eyes are covered with a hood. Also, an air-breathing animal deprived of its labyrinths, though a good swimmer, drowns if thrown into deep water, since it cannot orient itself by the sight of surrounding objects. Deaf mutes, whose labyrinths are very frequently un-

developed, though able to swim may become disoriented and drown if they fall or dive into deep water.

Righting reflexes may be demonstrated in the human subject. The baby of a few weeks old, for example, when lying prone raises the head into a nearly vertical position. When blindfolded and held by the pelvis in different positions in the air the head is moved toward the normal position.

LOCAL AND SEGMENTAL STATIC REACTIONS SUPPORTING REACTIONS

Magnus speaks of the simultaneous reflex contractions of both extensor and flexor muscles and other opposing muscles whereby the joints are fixed and the limbs converted into rigid pillars for the support of the body against gravity as the *positive supporting reaction*. This reaction is initiated by —

(a) Impulses discharged from the proprioceptors of the flexor muscles of the terminal segments of the limbs—digits and ankle or wrist, the pressure of an animal's paw upon the ground by stretching these muscles provides the adequate stimulus which calls forth simultaneous reflex contractions of the flexors and extensors of the knee (or elbow).

(b) Myotatic reflexes set up in the flexors of ankle and toes (plantar flexors), and of the corresponding forelimb joints, excessive extension at these joints is thus counteracted. Any tendency toward over-extension at the knee or elbow is also provided against through the reflex set up when the flexors of these joints are stretched. Similarly, any tendency of the knee or elbow to bend under the weight of the body calls forth a myotatic reflex from the extensors, which prevents any weakening of the supporting action of the limb.

(c) Impulses set up in the pressure receptors in the deeper layers of the skin of the sole when in contact with the ground, thus exteroceptive reflexes reinforce those of proprioceptive origin. The exteroceptor element is well shown in a decerebellated dog. When such a preparation is placed upon its back and the head strongly flexed, the hind limbs are flexed in all joints. Light pressure with the finger upon the toe pad then causes an extension of the limb, and if the finger be moved with the limb as it extends so that only very light pressure upon the pad is maintained, one has the sensation of the limb being drawn out by the finger. For this reason the movement has been called the "magnet reaction."

The relaxation of the muscles and the unfixing

of the joints which enables the limb to be flexed and moved to a new position is called the *negative supporting reaction*. It is brought about by raising the pad off the ground and plantar flexing the toes and ankle. The exteroceptive stimulus and the stretch stimulus to the *plantar flexors* are thus removed. The reflex "unlocking" of the limb is not, however, simply due to the removal of these stimuli, but has in addition a positive element, namely, the stimulus provided by plantar flexion and the consequent stretching of the dorsi-flexors of the toes and ankle—relaxation of the extensors of the knee or elbow and contraction of the flexors result.

The supporting reactions, though seen best in a decerebellate animal, can also be demonstrated in the decorticate preparation or in one whose brain-stem has been divided above the medulla oblongata. Segmental static reactions, e.g., flexion reflex and the crossed extension reflex have been described in chapter 64.

✓STATO-TONIC OR ATTITUDINAL REFLEXES

(1) *Tonic labyrinthine and neck reflexes acting upon the limbs*. These reflexes influence the tone of the skeletal muscles and thereby maintain the different parts of the body in an attitude appropriate to a given position of the head. They are investigated best in the decerebrate animal (p 965), the righting reflexes being then largely abolished. The proprioceptors concerned are in (a) the *labyrinth* (tonic labyrinthine (utricle) reflexes) which are brought into play by alterations in the position of the head in relation to the dimensions of space, (b) the *neck muscles* (tonic neck reflexes) which come into action when the position of the head is altered relatively to the body. In order to study separately the part played by each of these reflexes in any given reaction, the following procedures are adopted.

A. To exclude the neck reflexes

(a) Immobilization of the neck of an animal by means of a plaster of Paris bandage in order to prevent movement of the head in relation to the trunk. Any tonic effects resulting from a change in the animal's position must then be due solely to alterations in the position of the head in space (labyrinthine reflexes), or (b) section of the posterior roots of the first three or four pairs of cervical nerves.

B To exclude tonic labyrinthine reflexes

(a) Fixation of the head alone in some suitable apparatus. The tonic effect resulting from movement

of the body must then be due to an alteration in the position of the body in relation to the head, i.e., to movements of the neck, or (b) destruction of the labyrinths or section of the 8th nerves below the vestibular nuclei.

The labyrinthine reflexes exert an influence upon the tone of the extensor muscles which is in the *same* direction (increase or decrease) in all four limbs. The influence of the neck reflexes, on the other hand, is usually in *opposite* directions in the fore and hind pairs of limbs. The greatest degree of extensor tone is exerted through the labyrinth mechanism when the animal is supine and the mouth cleft inclined at an angle of 45° (fig 65 8).

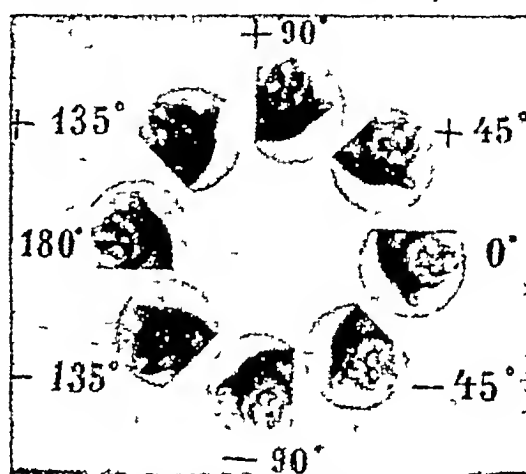


FIG 65 8. Diagrammatic representation of the positions of an animal's head, each marked with the angle which the mouth cleft makes with the horizontal plane (after Magnus)

above the horizontal plane, and is least when the mouth cleft makes an angle of 45° below the horizontal plane.

Extensor tone diminishes as the angle increases, it is minimal in the prone position with the mouth cleft at an angle of 45° below the horizontal plane.

In an animal on all fours the labyrinthine effect is therefore to increase or decrease the extensor tone in the muscles of all four limbs when the head is strongly extended or flexed, respectively. In the decerebrate labyrinthectomized animal, the neck reflexes alone operating, flexion of the forelimbs and extension of the hind limbs occur when the neck is flexed toward the sternum (ventriflextion) (see fig 65 4, p 965). Extension of the neck (dorsiflextion) produces the converse picture, i.e., extension of the forelimbs and flexion of the hind limbs. When therefore the neck is ventriflexted in the decerebrate animal with intact labyrinths,

the neck reflexes reinforce the tonic labyrinthine effect upon the forelimbs but antagonize that upon the hind limbs, the usual result is relaxation of the forelimbs with strong extension of the hind limbs. When the neck is extended the neck reflexes reinforce the labyrinthine effect upon the tone of the fore limbs but antagonize that upon the hind limbs. The effect of the neck reflexes upon the extensor tone of the latter again predominates, the extension of the forelimbs is maintained but definite relaxation of the hind limbs occurs.

Rotation of an animal's head (turning in the frontal plane of the skull) causes increased extensor tone of the fore and hind limbs on the side of the body toward which the jaw is rotated (*jaw limbs*)⁶ and reduces the extensor tone of the oppo-

ened in order to support the body's weight. A cat looking upwards to a bird in a tree extends the forelimbs and flexes the hind limbs, thus giving the back a suitable inclination which improves the position of the head and eyes, and places the body in a position preparatory for a spring (fig 659). When looking into a hole or beneath a cupboard the flexion of the forelimbs and extension of the hind quarters gives an opposite but no less advantageous inclination to the body. Depression of the back in the region of the last cervical vertebra brings the animal into a crouching attitude.

Stato tonic effects may be demonstrated in certain nervous lesions associated with a state analogous to decerebrate rigidity of animals (p 966). Turning the head to one side, for example, causes

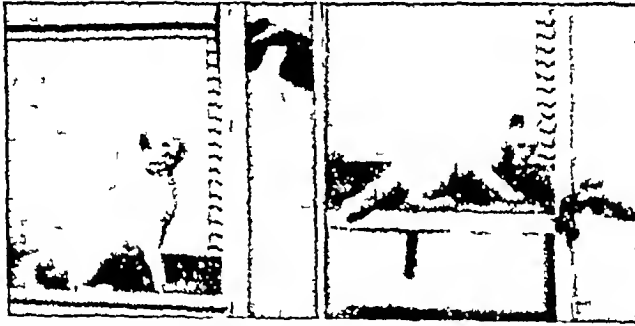


FIG 659 Photographs of a normal cat, showing the animal's posture (on left) when its attention is attracted by an object placed above it. Photograph of the same animal (on right) when its attention is drawn to an object below it. The difference between the two positions of the forelimbs is very marked, because in them the neck and labyrinthine reflexes reinforce one another, the hindlimbs are in much the same position in both cases, since the two sets of reflexes cancel one another (after Magnus).

site limbs (*skull limbs*)⁶. Inclination of the head toward one shoulder (lateral flexion) as when an animal turns a corner is accompanied by similar effects—extension of the limbs on the side of the body toward which the jaw (or snout) is inclined (*jaw limbs*) and flexion of the limbs on the opposite side (*skull limbs*).

Pressure upon the last cervical vertebra reduces the tone in all four limbs (*vertebra prominens reflex*).

The significance of these reflexes and their importance in the coördination of the postural muscles may be realized when the attitudes of the intact animal are observed. Thus when an animal turns to one side the limbs of that side are stiff-

an increase in tone of the extensors of the jaw limbs and hypotonicity in the limb muscles of the opposite side.⁷ When the head is in the position for maximal labyrinthine tone, i.e., when the patient is supine, and the neck extended the extensor tone of the paralyzed limbs increases, but becomes reduced in the prone position.

(2) *Tonic labyrinthine and cervical reflexes acting upon the eyes*. Tonic effects upon the eye muscles, analogous to those described for the skeletal muscles, result from changes in the position of the head. Labyrinthine and neck reflexes are responsible. Alteration in the position of the head with neck immobilized, or movement of the head in relation to the body after labyrinthectomy, is followed by compensatory eye movements. Turning the head downwards causes an

⁶ Magnus has introduced the term *jaw limbs* to indicate the limbs toward which the chin of man or the jaw of animals is rotated or inclined. The opposite limbs, i.e., the limbs to which the vertex of the skull is rotated, are called *skull limbs*.

⁷ Thus, according to Magnus, can also be shown in a certain percentage of normal infants, and in hydrocephalus it may be well marked.

upward movement of the eyes which are held in this position so long as the head position is maintained, the tone of the superior recti and inferior oblique being increased while that of the inferior recti and superior oblique is reduced. A corresponding compensatory movement of the eyes occurs when the head is turned upward. Similarly, when the head is turned to one side the internal and external recti of the two eyes cooperate to deviate the eyes outward or inward in relation to the head. Briefly, the eyes are moved in a direction opposite to that taken by the head, thus their original positions in space are maintained and the visual field existing prior to the head movement remains unaltered. It should be pointed out that the actual *movement* of the eyes is a stato-kinetic reflex (p 974) and due to a different mechanism (semicircular canals) from that which maintains the eye *position* while the head is held in the altered attitude. The latter is a statotonic reflex, dependent upon the utricle.

SUMMARY OF THE CENTERS FOR GENERAL STATIC REFLEXES

Magnus found that all the static reactions mentioned in the foregoing section could be obtained unaltered after removal of the cerebellum, as could also most of the stato-kinetic reflexes to be described in the next section. Nevertheless, this rather surprising fact does not necessarily imply that in the normal intact animal the cerebellum plays no part in these reactions. A contrary conclusion must be drawn from anatomical considerations and from observations in cerebellar disease in which disturbances referable to the labyrinth (e.g., abnormal positions of head, falling, etc.) are manifest. Moreover, Camus has demonstrated the arrival of action currents in the cerebellar nuclei during experimental stimulation (mechanical) of the labyrinth.

(a) *The righting reflexes*. All the labyrinthine righting reflexes, as well as the body righting reflexes acting upon the head, have their centers in the midbrain. According to Rademaker the labyrinthine righting reflexes are dependent upon the red nucleus. Destruction of the red nuclei or section of the decussation of Forel (p 1017) is claimed by this observer to abolish the righting reflexes.⁸ The center for the optical righting

reflexes is cortical, that for the body righting reflexes acting upon the head have their centers in the neighborhood of but not in the red nucleus itself (Magnus).

(b) *The stato-tonic labyrinthine reflexes* acting upon the skeletal muscles have their center in the vestibular nuclei.

(c) *The neck reflexes* are centered in the upper two or three cervical segments of the cord (see fig 65 10).

(d) The centers for the *tonic labyrinthine and cervical reflexes acting upon the eyes* are situated between the vestibular nuclei and the oculomotor nuclei.

Placing and hopping reactions. These were first studied by Rademaker and have been more recently investigated by Bard. Bard and Brooks describe five *placing reactions*.

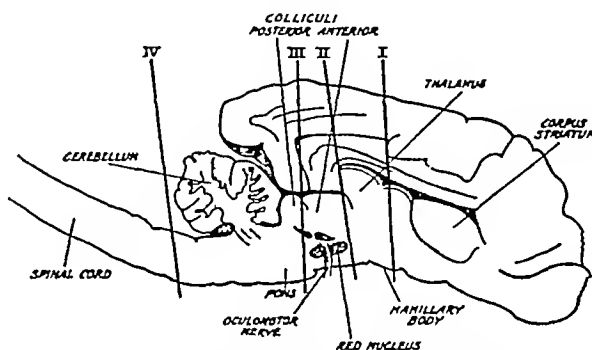


FIG 65 10 Diagram to illustrate the effects of sections through cat's brain at various levels. Line I, thalamic animal, righting, tonic labyrinthine and neck reflexes retained, little disturbance of muscular tone, Lines II and III, decerebrate rigidity, Line IV, behind vestibular nucleus, decapitate animal, extensor rigidity abolished, tonic neck reflexes retained. Section at level of second or third cervical segment of the cord—spinal animal—abolishes the tonic neck reflexes (after Magnus), III indicates Sherrington's original section.

(1) If a cat is held in mid-air with legs dependent and chin held up so that it cannot see anything below or in front, contact of the backs of the forepaws with the table's edge is followed by a quick movement of the limbs which brings the paws, soles down, precisely upon the surface of the table. (2) If the forelimbs of a cat are held down while the chin is brought in contact with the table near its edge, the forepaws when released are instantly raised and placed upon the table beside the chin, a movement which is usually followed by extension of the limbs and the assumption of a standing position. (3) If the fore- or hindlimbs of a cat standing or sitting upon a table are pushed over the table's edge, they are immediately lifted and placed in their original positions. (4) If one abducts, without holding, the limb of a standing cat, it is instantly returned to its previous position. (5) If a blind-folded cat is suspended in the air with forelimbs free, and its head brought toward some obstacle, at the instant that the vibrissae come into contact with the object the forepaws are raised and

⁸ The conclusion that the red nucleus is the essential center for the labyrinthine righting reflexes has been recently contested. Keller and Hare found that sectioning the rubrospinal tracts did not abolish them (See also Ranson and associates).

accurately planted upon the surface of the object. The first three of these reactions are due to stimulation of receptors upon the body surface (exteroceptors) and probably also of proprioceptors in muscles and tendons. The fourth is a purely proprioceptive reflex, the last is initiated from tactile receptors. The *hopping reactions* consist of limb movements which serve to maintain the standing position against any force acting upon it in a horizontal plane. When, for example, a cat is held so that its body is supported upon one fore- or hindlimb and is then pushed in one or other direction, the supporting limb hops quickly in the direction of the displacement, the foot being kept directly under the corresponding shoulder or hip. The hopping reactions are probably dependent upon myotatic (stretch) reflexes (p. 960).

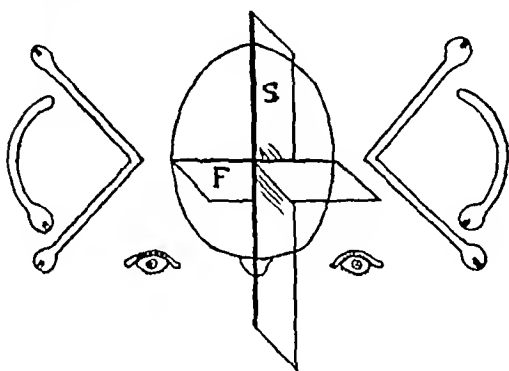


FIG. 65 11 Diagram showing the semicircular canals and their relation to the planes of the skull. S, median sagittal plane, F, transverse frontal plane (see text).

The placing and hopping reactions are controlled from the sensorimotor area of the cerebral cortex. Removal of this region by Bard and Brooks on both sides was found to abolish the placing reactions, and to produce an extreme degree of deficiency in the hopping reactions. Decortication (complete removal of the neocortex) produced no greater deficiency. The control exerted by the cerebral cortex is entirely contralateral, i.e., the component movements of the reactions on one side of the body are governed solely by the opposite side of the brain.

KINETIC, STATO-KINETIC OR ACCELERATOR REFLEXES

The term kinetic is attached to these labyrinthine reflexes because they are caused by the *movement* of the head, but since it is not the movement itself, but *acceleration* above a certain rate, angular (rotary) or linear (progressive), that is the adequate stimulus, *acceleratory reflexes* is a better name.

ANATOMY OF THE LABYRINTH

Before a description of other labyrinthine functions, e.g., the statokinetic reflexes, can be undertaken, a brief description of the structure of the labyrinth must be given. This structure consists of an auditory (cochlear) and a non-auditory portion. We are concerned here only with the latter which we shall refer to simply as the labyrinth. The *bony labyrinth* comprises a series of cavities tunnelled in the petrous part of the temporal bone. The cavities are the three semicircular canals, each of which opens by its two extremities into an ovoid chamber known as the *vestibule*. The bony labyrinth lodges a series of hollow membranous structures—the *membranous labyrinth*. The membranous labyrinth consists of (a) *Three semicircular canals* lying in the corresponding bony canals. (b) *Two sacs*, the *utricle* and *sacculle* situated in the vestibules (fig. 65 12).



FIG. 65 12 Diagram of the labyrinth. C, cochlea, CR, canals reunians, S, sacculle, U, utricle. Receptor areas indicated in heavy black.

THE SEMICIRCULAR CANALS One extremity of each canal shows an expansion known as the *ampulla* wherein is situated the specific sense organ. The membranous labyrinth is filled with fluid—the *endolymph*; a similar fluid—the *perilymph*—lies between its walls and the walls of the bony labyrinth. The canals lie in planes approximately at right angles to one another and are called respectively *external* or *lateral* (horizontal), *anterior* or *superior* (vertical) and *posterior* (vertical) (see fig. 65 11). It will be noted that the non-expanded extremities of the vertical canals join to form a common stem through which they communicate with the utricle (figs. 65 11, 65 12 and 65 13).

The external canal is directed with its convexity outwards and backwards. When the head is in the erect position this canal is only approximately horizontal, it is inclined backwards and downwards at an angle of about 30° to the horizontal plane. The vertical (anterior and posterior) canals both make an angle of about 45° with the frontal and the sagittal planes of the skull.* The anterior canal of one ear is

* That is, their planes cross both the frontal and sagittal planes diagonally.

therefore nearly in the same plane as the posterior canal of the other ear, whereas the posterior or anterior canal of one ear is at right angles to its fellow of the opposite side as well as to the other two canals of the same ear. The external (horizontal) canals of the two sides lie in the same plane.

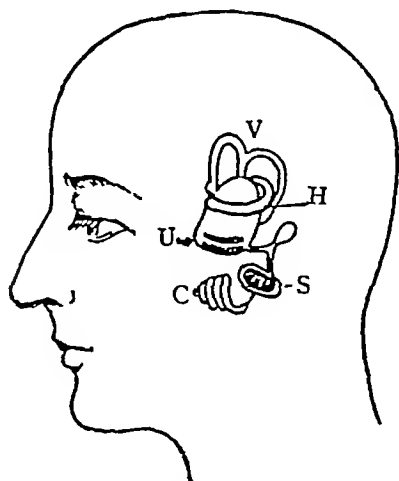


FIG 65 13 Diagram (not to scale) giving a lateral view of the internal ear within the skull. V, vertical canals, H, horizontal canal, U, utricle, S, saccule, C, cochlea (redrawn from Quix)

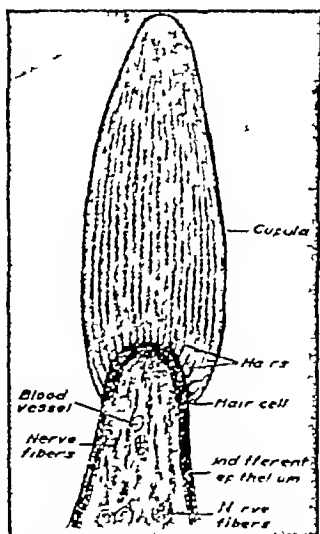


FIG 65 14 Drawing of crista. From Bailey's *Histology*, after Schafer

THE CRISTA. This is the receptor organ of a semicircular canal. It is situated in the ampulla and consists of a mound of sensory hair cells between which are non-sensory supporting cells. The sensory cells are surmounted by a gelatinous dome-shaped structure—the *cupula*, this contains fine longitudinal channels in which the hairs or cilia are lodged (fig 65 14)

THE OTOLITH ORGANS—THE UTRICLE AND SACCULE

In the wall of the *utricle* lie the openings of the semicircular canals. The utricle communicates through a fine canal with the cochlea and by a second small channel—the *ductus utriculosaccularis*—with the saccule. The sense organs of the utricle and saccule are called *maculae*. Each macula is a plaque of sensitive hair-cells covered by a gelatinous material—the *otolith membrane*—containing crystals or concretions of a calcareous substance—the *otoliths*. The macula of the utricle is situated on its anterior and medial walls, the two portions being joined at an angle of 140° . The sensitive hair cells are directed laterally and posteriorly. The saccule communicates with the cochlea through the *canalis reuniens*. The saccular maculae are situated obliquely, forming an angle of 27° with the vertical plane and 22° with the horizontal. The saccular maculae of the two sides incline towards each other so that if continued backwards they would meet behind the head. The dorsal lobe of the saccule also contains a macula whose hair-cells face downwards. The *sacculus* communicates with the utricle on the one hand and with the cochlea on the other. Its communication with the latter is through the duct of Hensen. The sense organ of the saccule is also given the name of *macula*¹⁰ and is constructed upon a plan similar to that of the utricle.

THE NERVOUS CONNECTIONS OF THE LABYRINTH

The central vestibular fibers are very widely distributed throughout the central nervous system, and are far from being fully known, but the following connections are fairly well established.

The impulses from the proprioceptors of the different parts of the non-auditory labyrinth are conveyed to the medulla by the vestibular branch of the 8th nerve. The cell bodies of the vestibular fibers lie in Scarpa's ganglion situated at the bottom of the internal auditory meatus. These cells are bipolar. Their afferent processes (dendrites) are distributed to the sense organs of the utricle, saccule and semicircular canals, and terminate in the hair cells. Their axons pass for the most part to the *superior, lateral, medial and inferior vestibular nuclei* in the medulla (a few go directly to the cerebellum) from which relay fibers follow three pathways (fig 65 15). Thus there are

(a) Ascending fibers from the superior vestibular nucleus which join the medial longitudinal bundle of the same side. From the medial vestibular nucleus impulses are relayed to the medial longitudinal fasciculus of the opposite side, from the inferior vestibular nucleus fibers ascend in the medial longitudinal fasciculus of both sides. The fibers from these nuclei terminate

¹⁰ The term "lapillus" is sometimes applied to the otolithic organ of the utricle while that of the saccule is spoken of as the "sagitta".

in the nuclei of the 3rd, 4th and 6th cranial nerves and constitute the vestibulo-ocular tract. Through these connections reflex movements of the eyes are brought about.

(b) Descending fibers, from the lateral (Dieter's) and the inferior vestibular nuclei. Those from the lateral nucleus constitute the vestibulospinal tract (ch

brum (temporal lobe) give rise to the sensation of vertigo.

As a result of his clinical findings I. H. Jones concludes that the fibers from the vertical and from the horizontal canals follow different paths. Those from the vertical canals do not, according to this observer, enter the vestibular nuclei but ascend to the pons where

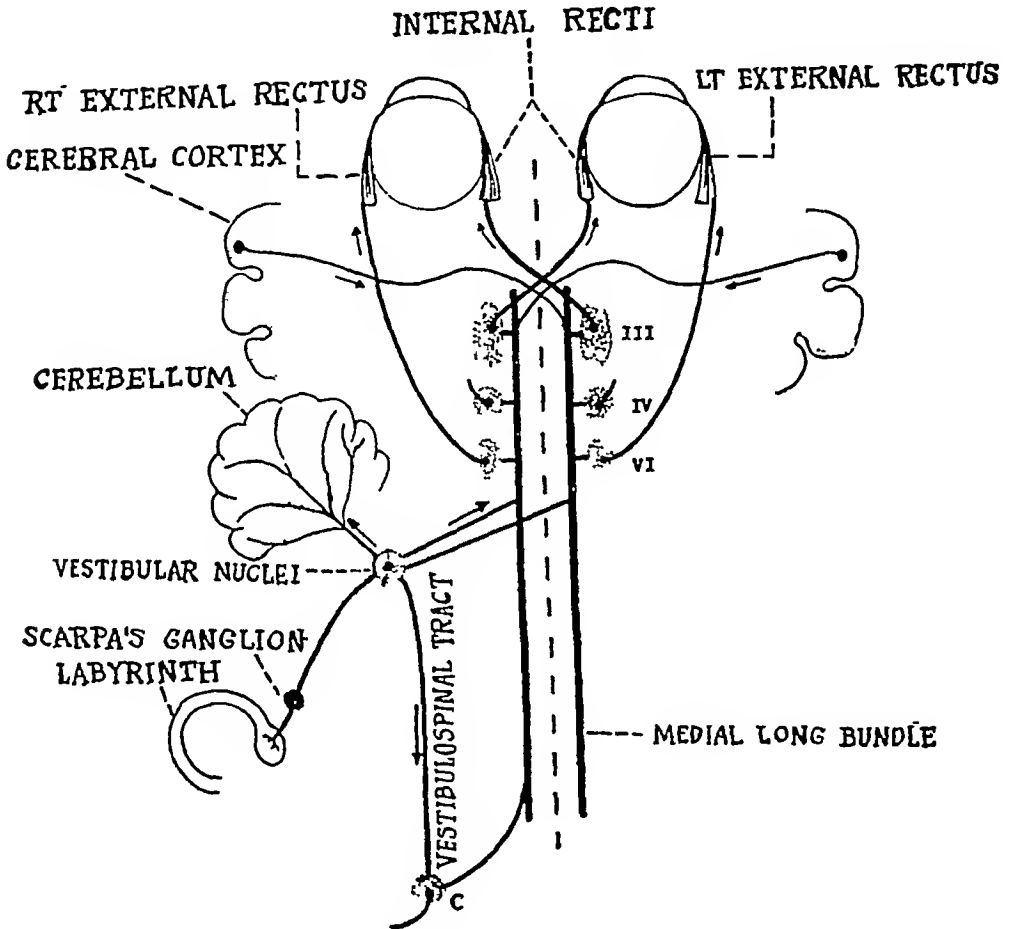


FIG. 65.15. Diagram of the principal vestibular pathways (see text). C, cervical segments of the spinal cord.

66), while those from the inferior nucleus descend in the spinal extension of the medial longitudinal fasciculus as far as the upper cervical region of the cord, where they synapse with the anterior horn cells. Through these connections impulses are conveyed to the skeletal muscles.

(c) Fibers pass via the inferior cerebellar peduncle mainly to the flocculonodular lobe of the cerebellum on the same side, but also to the fastigial nucleus of the same side. These constitute the vestibulo-cerebellar tract. Impulses conveyed along this pathway and thence through the superior cerebellar peduncle to the cere

brum (temporal lobe) give rise to the sensation of vertigo. As a result of his clinical findings I. H. Jones concludes that the fibers from the vertical and from the horizontal canals follow different paths. Those from the vertical canals do not, according to this observer, enter the vestibular nuclei but ascend to the pons where

EFFECTS FOLLOWING LABYRINTHECTOMY (IN THE DOG)

Unilateral labyrinthectomy (1) *Ocular* Both eyes are turned toward the side of the operation. The eye of this side also shows a downward deviation while the eye of the sound side deviates upwards.

(skew deviation) *Horizontal nystagmus* (p 979) with the slow movement toward the operated side

(2) *Lateral flexion and rotation of the occiput* to the operated side together with flexion of the thorax on the pelvis toward this side

(3) *The extensor tone of the limb muscles* is greater on the sound side than on the operated side, the limbs of this side being flexed and adducted, whereas those of the opposite side are extended and abducted

(4) *Spontaneous movements* These are all toward the operated side and consist of (a) circling, i e, turning of the body around a vertical axis, (b) rolling, i e, rotation around a horizontal axis, in the rabbit the rolling movements are very violent, (c) side to side movements of the head (head nystagmus), (d) falling to operated side, (e) stepping gait

Many of the foregoing features, e g, nystagmus, rolling and circling movements, are *irritative* in nature, i e, the result of irritation of nerve endings by the operative trauma. They therefore improve with time. The others, e g, oscillating movements of the head, head torsion, stepping gait, falling and asymmetrical distribution of tone, are deficiency phenomena.¹¹ These, for the most part, also improve since they tend to be compensated by visual and other non-labyrinthine reflexes, especially in higher animals, who also learn to exercise cerebral control over the abnormal muscular activity

Bilateral labyrinthectomy produces irritative effects similar to those described above, the direction of the movements being variable and dependent upon which ear shows the greatest degree of irritation. The labyrinthine static reflexes are, of course, abolished but in the more intelligent animals the defects of orientation are largely compensated for by visual reactions. The normal reactions to rotation (p 979) are abolished

LABYRINTHINE FUNCTION

The semicircular canals are organs of dynamic sense, the otolithic organs subserve both dynamic and static senses. The hair cells of the cristae in the canals respond to angular acceleration (rotation), and are therefore called *rotary receptors*. The hair cells of the macula on the medial wall of the

utricle, and probably the macula of the saccule, as well, are stimulated mainly by changes in position of the head, and by linear acceleration. They are called *gravity receptors*.¹²

The semicircular canals The rotary receptors of the cristae respond and give rise to a sensation at the commencement and at the cessation of a rapid rotary movement. Though impulses at low frequency are discharged along the nerve fibers while the head is at rest, a sharp increase in impulse frequency occurs at the commencement of the rotary movement, and at its end. The rapid discharge lasts for about 25 seconds. The rotary receptors, therefore, give no information of a continuous rotary motion of unvarying speed, though any alteration in speed of rotation (angular acceleration) arouses a sensation. The reason for this will be evident upon consideration of the stimulating mechanism.

Stimulation of the sensitive hair cells of the cristae is brought about through the effect which the movement of the head exerts upon the endolymph within the canal, the particular canal which is stimulated depending upon the plane in which the rotation occurs. The canals are of capillary diameter (0.1 to 0.2 mm in man), and the endolymph is of relatively high viscosity (2 to 3 times that of water). The membranous labyrinth is a practically closed system and is supported on the outside by perilymph, and the unyielding walls of the osseous labyrinth.

When the head turns quickly from a position of rest, say to the right (see fig 65.17 and table 91), i e, in a clockwise direction, and in the plane of the horizontal canals (around a vertical axis), the endolymph within the canal, owing to its inertia, does not move at first with the walls of the canal. This lag is equivalent to a movement or flow of endolymph to the left—anticlockwise—in both canals. But owing to the relative positions of the two ampullae such a flow of endolymph causes the cupula—the gelatinous mass surmounting the crista, and in which the hair cells are embedded—to swing toward the utricle in the right ampulla and away from the utricle in the left ampulla. Deflection of the cupula causes, in turn, deformation of the hair cells which acts as a stimulus when the swing is toward the utricle (right) but is “inhibitory” if it is away from the utricle (left).

¹¹ Cocaine injected into the labyrinth produces the deficiency phenomena characteristic of labyrinthectomy, but when applied to the nerve stump after removal of the labyrinth it prevents the occurrence of the irritative phenomena.

¹² Breuer (1874–1891) was the first to point out the difference in function between the semicircular canals and the utricle.

These events are reversed when the rotation is to the left—anticlockwise

Thus, a clockwise rotary movement of the head (e.g. a quick turn) around a vertical axis stimulates the right horizontal canal, an anti-clockwise motion excites the left, or, put in another way, the rotary receptors are stimulated only when the narrow part of the canal is leading and the ampulla trailing

The receptors of the vertical canals, on the other hand, are stimulated only when the ampulla is leading and the narrow part of the canal bringing up the rear. There is no explanation for this difference between the horizontal and vertical canals

The maximal response is given by a canal when

TABLE 91

SEMICIRCULAR CANAL	ROTATION ABOUT THE					
	Longitudinal axis		Transverse axis		Vertical axis	
	Right	Left	Forwards	Backwards	Clockwise	Anticlockwise
Right ant. vert.	●	⊗	●	⊗	⊗	●
Left ant. vert.	⊗	●	●	⊗	●	⊗
Right post. vert.	●	⊗	⊗	●	⊗	●
Left post. vert.	⊗	●	⊗	●	●	⊗
Right horizontal	○	○	○	○	●	⊗
Left horizontal	○	○	○	○	⊗	●

Responses of the six semicircular canals (of fish, the thornback ray) to rotation ● excited, ⊗ inhibited, ○ unaffected ant. vert. = anterior vertical, post. vert. = posterior vertical (after Löwenstein and Sand)

it is rotated in its own plane¹² No response is aroused from a horizontal canal by rotation in a plane at right angles to this. But a vertical canal, though most effectively stimulated by rotation in its own plane (diagonal) also gives a lesser response when rotated in other planes. Thus a quick tilt of the head, forward or backward (rotation around a transverse axis), or laterally toward a shoulder (rotation around an anteroposterior axis) stimulates the vertical canals. The threshold of excita-

¹² It will be recalled that the vertical canals are placed diagonally in the head, and that an anterior vertical canal is in the same plane with the posterior canal of the opposite side, therefore, in a manner analogous to that of the horizontal canals, a diagonal rotary movement forward will maximally stimulate an anterior canal and suppress the activity in the opposite posterior canal. A backward rotary movement would act in the opposite way

tion of a horizontal canal is an angular acceleration between 2° and 3° per sec.

As any rotary movement is continued for longer than a few seconds, and the inertia of the endolymph is overcome, the cupula regains its resting (unbent) position, at the cessation of the movement the momentum of the fluid causes bending of the cupula in a direction opposite to that at the commencement of the movement. This is why no sensation occurs, during prolonged rotation, but a sensation of turning in the opposite direction is felt when the rotation suddenly ceases

The mechanism of the semicircular canals is admirably suited to signal rapid turns of the head, i.e., a rotary movement in one or other plane. But it is poorly adapted to report continuous rotation at a constant speed (for, as Adrian points out, this is a movement to which the body throughout its life is rarely if ever subjected, and has therefore not developed suitable receptors for recording it). Continuous rotation tends to confuse rather than to inform, and also, because of the swing of the cupula in the reverse direction when the rotary movement ceases, an illusion of rotation (in the opposite direction) is created. The ballet dancer avoids the confusion in a pirouette by making a series of alternating acceleration and deceleration.

A brief outline of the experiments on vestibular function. The foregoing account of vestibular mechanisms is based upon the work of many investigators. Some of the more recent experiments will be briefly outlined.

McNally and Tait studied the effect on frogs of tilting and rotation after denervation of one or more canals in various combinations, and after destruction of the utricle or saccule. Steinhausen elucidated especially the movements of the cupula in the canals of the pike whose labyrinth can be visualized in the living fish after the introduction of India ink, the cupula showing up clearly against the darker background. This structure was shown to completely partition the cavity of the ampulla, being in contact everywhere around its free circumference with the membranous wall. It was shown to move during caloric stimulation as well as during angular acceleration, but not during electrical stimulation. Later, similar studies were made by Dohlman which confirmed Steinhausen's findings. Dohlman followed the endolymph movements by means of a drop of oil introduced into a canal. Ross employed the right half of a frog's head held in an apparatus, whereby the labyrinth could

be turned and fixed in any desired plane, and subjected to rotation. The electrical changes were recorded from a filament of the auditory nerve. Clockwise rotation caused a response from the horizontal canal, whereas an anticlockwise movement was ineffective. Steinhausen's conception of a deflected cupula caused by an endolymph movement was verified. Lowenstein and Sand employed the surviving isolated labyrinths of the pike and ray (*Raja clavata*) and recorded the electrical potentials from individual canals and found that the nerve endings of the cristae discharged impulses continuously during rest at a low rate but at increased frequency during angular acceleration and deceleration. Adrian, experimenting with cats, recorded the potentials from single units by means of a fine wire electrode inserted into the medulla in close proximity to the vestibular nuclei. The main findings of others in cold blood animals were confirmed.

EFFECTS OF STIMULATION OF THE SEMICIRCULAR CANALS

The semicircular canals respond to any one of the following forms of stimulation: (1) *Rotation* (angular acceleration) of the head on a vertical, transverse or anteroposterior axis; (2) *Caloric*—syringing the ear with a hot or cold solution; (3) *Galvanic*; (4) *Mechanical*.

Rotation

The effects produced by rotation are: (a) *eye movements*, (b) *vertigo*, (c) *reactions of the neck and limb muscles*, (d) *reactions of the autonomic nervous system*.

(a) *Eye movements*. When an animal's head is turned sharply on a vertical axis, i.e., in the plane of the horizontal canals, the eyes deviate in the opposite direction, and then return with a quick jerk to the original forward-looking position. The movement causing the slow deviation of the eyes is really a rotary one through a small arc.

If the horizontal rotation is continuous, a rhythmic to and fro movement of the eyes occurs when the rotary movement ceases. This is called *nystagmus*. It can be easily demonstrated in man by rotating him with eyes closed, in a revolving chair, at a rate of about 10 complete turns in 20 seconds. In order to stimulate the canals maximally, the head should be inclined forward about 15° when the rotary movement is almost exactly in the plane of the horizontal canals. When the

rotation is suddenly stopped (angular deceleration), and the subject opens his eyes, the rapid to and fro motion of the eyes is seen. It consists of a slow deviation to one side, and a quick return to the normal position. This post-rotary nystagmus lasts for about 20 seconds in a normal person, its quick

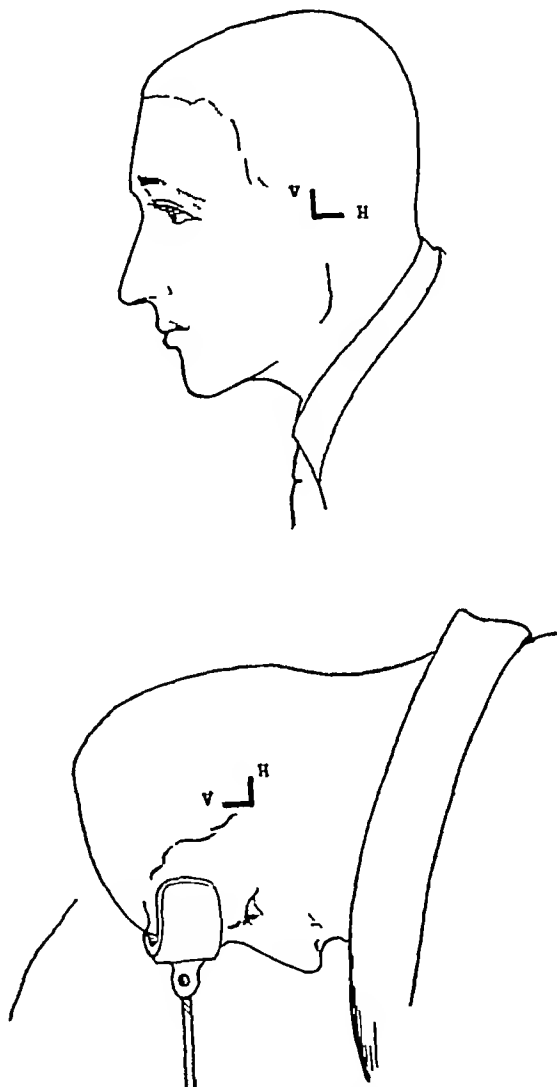


FIG 65 16 Positions during rotation. Upper, horizontal canals in plane of rotation. Lower, vertical canals in plane of rotation. V, vertical canals, H, horizontal canals.

component is in a direction opposite to that of the rotary movement. By convention *the quick movement or jerk to the normal position is taken to designate the direction of nystagmus from whatever cause*. Thus, a right horizontal post-rotary nystagmus is caused by rotation to the left. At the beginning of the rotation to the left the nystagmus would be in the opposite direction—quick movement to the right (figs 65 16 and 65 18).

There are two other forms of nystagmus—*vertical* and *rotary*—in which the eyes oscillate up and down, and around an anteroposterior axis, respectively. These types result from the simultaneous excitation of both vertical canals. Vertical nystagmus is induced when the canals are stimulated with the sagittal plane of the head in the plane of rotation, that is with the head inclined about 90° to one or other shoulder. The rotary type follows when the frontal plane of the head is bent forward 90° or more, or backward 30° to 60° . In each instance the post-rotary nystagmus is observed after the head has been brought upright.

Central connections responsible for nystagmus. The slow phase of nystagmus is initiated from the semi-

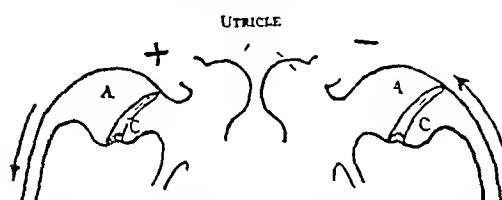


FIG 65 17 Diagram of ampullae (A) of horizontal canals and utricle, showing deflection of the cupula during rotation. Plus and minus signs indicate increase and decrease, respectively, in the impulse frequency from the hair cells, arrows indicate the direction of rotation, C, cupula

circular canals and has its center in the vestibular nuclei, from which impulses are discharged, in part at least, through the median longitudinal bundle to the eye muscles. The quick movement is entirely central, but the nervous pathways upon which it depends are uncertain. However, its neural mechanism must lie in the brain stem between and including the nuclei for the third nerves and the vestibuli nuclei, for nystagmus occurs after transections of the brain above and below these levels, respectively. It is not abolished by ablation of the cerebellum. Even the nuclei of the 3rd and 4th nerves may not be necessary, for in the rabbit nystagmus can be induced after section just above the abducent nucleus. Lorente de N6 has located the center for the rapid phase in the formatio reticularis in the region of the abducent nucleus. He also found that nystagmus could still be produced after section of both median longitudinal bundles. There may be a double pathway from the vestibular nucleus to the nuclei of the ocular nerves—through the medial longitudinal bundle and through the formatio reticularis.

(b) *Vertigo*. Vertigo consists of the familiar whirling sensation or giddiness, and the disturbances of equilibrium which follow spinning. When

the subject is rotated with head erect, his vertigo is in the horizontal plane, i.e., in the plane of the external canals but in the direction opposite to that of the previous rotation (sensation of counter-rotation).

If the subject is rotated with the head bowed forward at an angle of about 90° , in order to stimulate the vertical canals, and the head maintained in this position after rotation, the sensation is the same as if the head position had been upright during rotation, i.e., one of counter-rotation in the horizontal plane. If, however, the head is brought upright after rotation there is then a sensation of falling to one or the other side, i.e., of rotation of the body in the frontal plane, the sensation is of falling to the side away from that toward which the body had been rotated. Rotation with the head

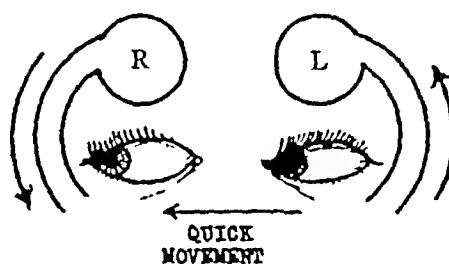


FIG 65 18 Postrotary nystagmus

bent backwards to an angle of 60° and then brought upright gives a sensation of falling to the same side as that toward which the body had been rotated (see table 92). A sensation of falling forward or backward (i.e., rotation in the sagittal plane) results from rotation with the head flexed on one shoulder and then brought upright.

Vertigo arises from other causes than rotation. It is a common symptom in disease, constitutional or neurological, and occurs in alcoholic intoxication, seasickness, swinging, etc. Also, just as the labyrinth influences the movements of the ocular muscles, so, conversely, labyrinthine function may be disturbed and vertigo produced as a result of some unusual or abnormal action of the eye muscles. Vertigo and its associated phenomena are, therefore, common effects of eye strain or of viewing the landscape from a moving train, an ocular element is also an important contributory factor in the causation of sea-sickness. For example, if one sees a moving picture, which if actually experienced would cause giddiness, a sensation of vertigo is aroused. But with whatever condition vertigo is associated, whether cardiovascular, renal, toxic, gastrointestinal or neurological, its immediate

cause is excitation of the semicircular canals or of their central connections

(c) *Reactions of the neck and limb muscles In animals* Stimulation of the horizontal canals by a rotary motion causes alterations in tone and movements on both sides of the body which tend to resist the rotary movement, and enable the animal to maintain its balance When, for example, a frog is rotated on a round table, say to the *left*, the tone

cites a simple means for their demonstration When a frog is rotated, say to the left, around its long axis by suspending it by a thread fixed to its jaw, and giving the thread a twist between the finger and thumb, the animal's right hind limb is extended and abducted, and the toes spread, while the left hind limb is bent at knee and ankle, and the toes flexed As the movement to the left comes to an end, and turning to the right begins the limb movements are reversed Similar movements may occur in the forelimbs

TABLE 92

POSITION OF HEAD	STIMULUS	TYPE AND DIRECTION OF NYSTAGMUS (RAPID PHASE)	SENSATION OF VERTIGO, HEAD UPRIGHT AFTER ROTATION	PAST POINTING	FALLING (AFTER BRINGING HEAD UPRIGHT)
Upright or forward	Rotation right	Horizontal, left	Turning left	To right	To right
Forward 120°	Rotation right	Rotary, left	Falling to left	To right	To right
Backward 30°-60°	Rotation right	Rotary, right	Falling to right	To left	To left
Inclined to right shoulder	Rotation right	Vertical, downward	Falling forward	Upward	Backward
Inclined to left shoulder	Rotation right	Vertical, upward	Falling backward	Downward	Forward
Backward 30°-60°	Caloric 112°F to ear or 68°F to left	Horizontal, right	Falling to right	To left	To left
Forward 90°-120°	Caloric 112°F to left ear or 68°F to left Galvanic	Horizontal, left Mixture of horizontal and rotary Direction same as that of current	Falling to left	To right	To right

With stimuli opposite in sign to those given in the table, the reactions are reversed

of the neck muscles is increased on the right side and reduced on the left, the head turning to the *right* The limbs, especially the hind limb, on the left are extended, those on the right are flexed the animal may move around to the right Cessation of the rotary motion causes a reversal of these effects upon the neck and limbs In the snake, lizard, and reptiles with very long necks, the head may deviate to one or other side by as much as 100° or more Sometimes a quick return to the normal position occurs, as in nystagmus

Dusser de Barenne, who first described these reflexes,

Corresponding effects follow rotation, e g, a quick tilt, in the plane of a vertical canal Tait and McNally described the respective vertical canals as affecting the limb muscles on one or other "corner" of the body Thus, stimulation of the right anterior vertical canal by a quick tilt diagonally forward and to the right causes contraction of the right forelimb stimulation of the left anterior vertical canal by tilting forward and to the left causes a corresponding movement of the left forelimb Backward tilting to right or left causes, respectively, contraction of the muscles of the right and of the left hind limbs

The past-pointing test of Barany Under ordinary circumstances a normal person if he places his finger upon a certain spot has no difficulty in hitting the mark again with his eyes closed. After rotation, though able to place his finger upon a mark with his eyes open, he cannot find it again when his eyes are closed. The finger *deviates* or *past points* to one or other side, or above or below the mark, the direction of the miss-aim being dependent upon the direction of the previous rotation and upon the position of the head during rotation (see table 92). Past-pointing is not reflex in nature but is a voluntary motor act, the error in judgment is the result of the associated subjective phenomenon of vertigo, a subconscious correction being made in the opposite direction for the false sensation. The deviation of the finger and the vertigo are therefore in opposite directions.

Other post-rotary reactions If the body is rotated in the plane of the horizontal canals and the movement stopped, the head (eyes closed) then turns in the direction of the rotation. If rotation is carried out with the head in one or other plane of the vertical canals and the head after rotation is brought upright while the eyes are closed, the body leans to one or other side, backward or forward, according to the position of the head during rotation. The subject may actually fall in the direction to which the body leans. The phenomenon is virtually a past-pointing of the entire body. The actual fall is, therefore, opposite in direction to the vertiginous sensation of falling. That is, the subject has the illusion that he is leaning to one side (i.e., a sense of being rotated in the frontal plane of the skull), and in order to correct his supposed false position and retain his balance, leans actually in the *opposite* direction, and therefore falls if he leans too far.

One particularly interesting reaction resulting from excitation of the semicircular canals is that which has been appropriately called the *disc throwing* or *discobolus* attitude (fig 65 19). Though occurring after rotation it is evoked most readily by caloric (especially cold) or galvanic stimulation of the canals. A stimulus applied, say to the left ear, causes a twisting of the thorax upon the pelvis and rotation of the head to the stimulated side. When the arms are raised they are also turned toward this side with the left limb lower than the right. After a short time the attitude is reversed, the body swings round and takes up a position in the opposite direction. The attitude may reverse its

direction several times. It is due to reflex alterations in tone of the musculature on the two sides of the body.

(d) *Reactions of the autonomic nervous system* Excitation of the semicircular canals in man is not uncommonly followed by nausea, vomiting and pallor. A fall in blood pressure of 10 mm. or so may occur, together with slowing of the heart by 8 or 10 beats per minute. In the rabbit, syringing the ear causes vasodilatation and a fall in blood pressure. During rotation the pupil constricts, pupillary dilatation occurs upon cessation of the rotary movement.



FIG 65 19 So-called 'discobolus' position resulting from caloric or galvanic excitation of the labyrinth (from Camis after Wodak and Fischer)

Camis has shown that after labyrinthectomy, the usual effect upon blood pressure of stimulating the central end of the vagus is reversed. For example, the effect of stimulating the central end of the vagus in the normal dog is frequently a rise in blood pressure accompanied by vasoconstriction of the vessels of the hind limb. After unilateral labyrinthectomy stimulation of the central end of the vagus on the operated side causes a fall in blood pressure together with a reduction of limb volume on the same side. Stimulation of the vagus in a bilaterally labyrinthectomized animal results in a fall in blood pressure and a reduction in volume of both limbs—a relationship between blood pressure and limb volume which is the reverse of the normal (p. 295). There is no satisfactory explanation of these paradoxical reactions. Camis suggests that labyrinthectomy abolishes the activity of the vasoconstrictor center. Stimulation of the central end of the vagus would then result in pure vasodilator effects, and a fall in blood pressure. The reduction in the volume of the limb may be simply a passive effect—the drainage of blood from its vessels. Bayliss found, for example, that after removal of the abdominal sympathetic (constrictors to the vessels of the abdomen, and hind limbs), a fall in blood pressure, accom-

panied by a reduction in limb volume resulted from stimulation of the central end of the vagus

(2) *Caloric stimulation*

The effects of caloric stimulation are similar to those following rotation but possess an advantage in that one or other ear can be examined separately. The ear to be tested is syringed with hot (112°F) or with cold (68°F) water.¹⁴ When the head is bent backward through 60° the horizontal canals are brought into a vertical position. The douche causes a greater change in the temperature of the endolymph in the part of the canal lying nearer to the external meatus than in the part more deeply situated. Convection currents are set up which stimulate the receptors of the cristae, horizontal nystagmus and vertigo result. The change in temperature in the canal follows the irrigation of the external meatus by about 3 seconds. The direction of the convection currents, which of course are due to changes in the specific gravity of the endolymph resulting from heating or cooling, is determined by the temperature of the douche fluid. Thus a cold douche causes currents away from the ampulla, a hot douche causes ones toward the ampulla (see table 92). Caloric stimulation of the vertical canals is effected by douching with the head upright.

(3) *Galvanic stimulation*

In the employment of this method of stimulating the canals electrodes are placed one upon each mastoid process or, more usually, one on a mastoid process and the other on some indifferent part of the body. The current required in normal persons is from 2 to 7 milliamperes. Owing to current spread all six canals are excited usually by electrical stimulation. The nystagmus has a rotary movement as well as a horizontal. The horizontal movement is with the current. Thus, when the cathode is over the right labyrinth the nystagmus is to the right, though the head may turn to the left. The resulting nystagmus is a mixture of the horizontal and rotary forms, its direction being the same as that of the current. Thus, when the cathode is over the right labyrinth, the nystagmus is to the right and vice versa. The effects occur only during the make and break of the current. In the absence of the labyrinth the galvanic current produces its effects by stimulation of the vestibular nerve. Consequently when, as the result of disease, labyrinthine function has been destroyed, this method affords a means of determining the condition of the nerve. Whereas rotary, caloric and direct mechanical stimulation involve essentially the same mechanism—movement of the cupula—the galvanic current is a direct electrical stimulation of the cristae. Steinhausen observed no movement of the cupula during this form of stimulation.

¹⁴ Douches much nearer the body temperature than these will stimulate

(4) *Direct mechanical stimulation of the canals in animals*

Ewald cemented a metal cylinder over a hole made in the bony wall of the horizontal canal. A piston fitted into the cylinder could be operated by air pressure. During the descent and ascent of the pneumatic hammer the membranous canal was compressed and decompressed respectively. The endolymph during compression and decompression moved, respectively, toward and away from the ampulla. An endolymph movement toward the ampulla caused a movement of the head and eyes to the opposite side. Decompression caused a weaker movement in the reverse direction.

The otolithic organs, utricle and saccule, gravity receptors. The role played by the utricle in the static reflexes has been dealt with in previous sections (see righting reflexes). But the gravity receptors respond also to linear acceleration, that is, to movements in a straight line forward, backward, up or down, or laterally, as well as to slow tilting of the head on its transverse or its anteroposterior axis. The gravity receptors are also stimulated by a sudden cessation of a linear movement (deceleration), or a sharp change in the speed of such a movement. The receptors are stimulated when the otoliths are hanging down, and thus exerting a gravitational pull upon the sensitive hairs. It is possible, that bending the hairs also acts as a stimulus. The utricular organs are also stimulated by centrifugal force, and under certain conditions by angular acceleration.

In the cat, when the head is level the receptors cause a discharge of impulses at a relatively low frequency (at about 6 per second). When the head is tilted laterally (on its anteroposterior axis—cheek down) from the level position to an angle of 20°, the impulse frequency increases up to 95 per second in the nerve on the side to which the head is tilted, but the resting discharge on the opposite side ceases. Thus, in a lateral tilt to the right the otoliths on that side (cheek down) are dependent and presumably exert a gravitational pull, whereas those on the opposite side are *resting* on the maculae, and are inactive. The greater the degree of downward tilt from the level position the higher is the impulse frequency (see fig. 65.20). If the head is held in the tilted position the impulse frequency declines slowly, for the gravity receptors are of the slow-adapting type. Tilting around a transverse axis (snout up or down) also evokes a response, though less readily than does a lateral tilt.

A pull upon the otoliths as a result of their inertia should also be caused by linear acceleration, i.e., a progressive or translatory movement in any of three directions, fore and aft, up or down (vertical linear acceleration),¹⁵ or sideways. Records of the action currents from the labyrinth show that this actually occurs, horizontal or vertical acceleration stimulates the gravity receptors in a lateral movement on the side from which the movement is made. Thus, a quick movement to the left stimulates the receptors on the right, owing to the inertia of the otoliths, acceleration

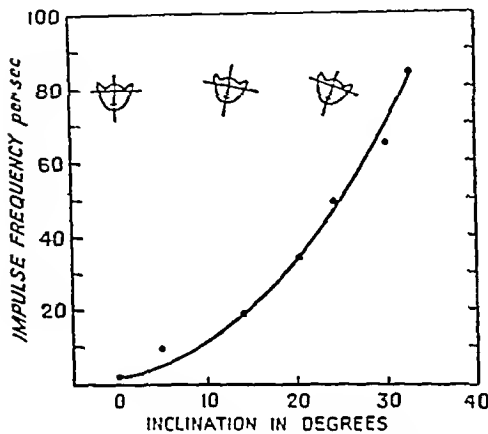


FIG 65.20 Response of gravity receptor (otolith organ), showing the relationship between degree of tilt of head and frequency of impulses in vestibular nerve. (After Adrian.)

laterally and to the left acting like a gravitational pull upon the receptors of the right side. Adrian found that rapid fore or aft, or up or down (vertical linear acceleration) also caused responses from the labyrinth. The threshold of excitation of the otolith organs by linear acceleration is around 12 cm per sec. per sec.

The saccule The role played by the saccule in the responses just described has not been settled unequivocally. Though anatomically this structure appears to be part of the vestibular mechanism it has been thought by many to have a cochlear rather than a vestibular function. McNally and Tait observed no loss of any known vestibular function after denervation of the

sacculi. The auditory nerve from a frog's labyrinth which has been denervated except for the saccule does not respond to vestibular stimuli by a discharge of impulses, but gives a response when a vibrating tuning fork is brought near it. Lorente de Nô, on the other hand, traced saccular fibers to the vestibular nucleus, but none could be found connecting this organ with the cochlear nucleus. Adrian considers that the sacculi rather than the utricles are more likely responsible for the responses caused by lateral tilting and for upward linear acceleration, while the utricles are excited by downward and horizontal accelerations and tilting around a transverse axis. A macula is situated on the medial wall of the saccule and faces laterally, a position suitable for stimulation of gravity receptors by a lateral tilt. Another macula lies in the roof of the saccule and faces downward and could be stimulated by upward linear acceleration.

In the guinea-pig the otolith membranes can be detached by centrifuging the animal (at a speed of 1000 meters per minute). All responses to linear acceleration or to different positions of the head are then lost, but those caused by angular acceleration are retained.

VESTIBULAR REACTIONS IN DISEASE

Abnormal vestibular reactions are seen in various diseased conditions involving (a) the labyrinth, (b) the vestibular nerve, or (c) the vestibular centers or central pathways (e.g., Deiter's nucleus, medial longitudinal bundle, cerebellum). Normally, some slight nystagmus may occur upon looking for a time to the right or left, but if nystagmus is present when looking forward or if pronounced when the eyes are turned to one side, it is pathological. A spontaneous vertical nystagmus suggests a lesion of the brain stem, it is not seen in disease of the labyrinth itself or of the vestibular nerve. The phenomenon of past-pointing, unless induced artificially, is always pathological and suggests a cerebellar lesion. Spontaneous vertigo and falling in one or other direction may result from disease of the cerebellum, the labyrinth or the pathways of the vestibular fibers. On the other hand, a lesion in one or other of these situations may cause a failure of the usual reactions following rotary or caloric stimulation. In deaf mutes, the labyrinthine reactions, as a rule, are absent. Again, the reactions to artificial stimulation may be abnormal, for example, in a lesion of the brain stem vertical nystagmus may occur in

¹⁵ In a quick linear movement forward or backward, or vertically, the gravity receptors would be in the same plane as that of the movement, which implies that they can respond to a pull acting in their own plane, though probably less effectively than to one acting at right angles. But in this case bending of the hairs rather than a pull upon them may be the effective stimulus.

response to a stimulus which normally causes nystagmus of the horizontal type

MÉNIÈRE'S SYNDROME

The features of this condition are paroxysmal attacks of vertigo, noises in the ears (tinnitus) and usually impairment of hearing of the inner ear type (ch 79) Loss of consciousness may occur There may be spontaneous horizontal nystagmus The sense

organs of the cochlea (organ of Corti) as well as those of the semicircular canals are affected in this disease, which leads Crowe to believe that a common factor, namely, alterations in pressure of the endolymph or in its chemical composition is the most likely cause of the symptoms The volume of the endolymph is increased The surgical treatment of the condition consists of severing the vestibular portion of the 8th nerve in the internal auditory meatus, leaving the cochlear division intact (McKenzie, Dandy)

CHAPTER 66

THE SPINAL CORD AND BRAIN STEM (MEDULLA, PONS AND MID-BRAIN)

OUTLINE OF THE INTERNAL STRUCTURE OF THE CORD

In figure 66 5 the spinal cord is shown in cross-section. The *gray matter*, centrally placed, is in the form roughly of an H or the two wings of a butterfly. It is composed of a mass of nerve cell bodies and nerve fibers (dendrons and axons), mostly unmyelinated, supported by a framework of neuroglia. The ventral and dorsal portions of each lateral half of the gray mass (i.e., each arm of the H) are commonly referred to, respectively, as the ventral (or anterior) and dorsal (or posterior) horns, but since the gray matter extends throughout the length of the cord "column" is a more suitable term than "horn". In the ventral columns are situated the large bodies (100 μ in diameter) of the motor neurons whose axons leave the cord by the ventral roots. Each axon ends in a group of skeletal muscle fibers—the neuromuscular structure constituting the so-called motor unit (p 954). In the thoracic and upper lumbar segments, the gray mass lying between the ventral and dorsal columns shows a small lateral projection. This is the lateral column or horn, it contains a cluster of nerve cells (the *intermedio lateral cell column*) they give origin to sympathetic (preganglionic) fibers which leave the cord by the anterior (ventral) nerve roots. The well-defined collection of cells occupying the inner part of the base of the posterior horn is known as *Clarke's column*, this group since it is confined almost entirely to the thoracic region of the cord is also known as the *dorsal nucleus*. It is homologous with the nucleus cuneatus of the medulla (p 990). To the outer side of the base of the posterior column is an area where strands of white matter and prolongations from the main mass of gray matter intermingle to form a delicate interlacement. This is known as the *reticular formation* (formatio reticularis) and is most prominent in the cervical region. It is continuous with the reticular formation of the medulla and pons. At the apex of the dorsal horn is a cap of gelatinous material containing groups of small nerve cells possessing many dendrites and named the *substantia gelatinosa of Rolando*. It is believed

to receive the finest fibers (myelinated and unmyelinated) of the posterior roots. The remaining part of the substance of the dorsal horns constitutes the *chief sensory nucleus*. A canal pierces the bar or isthmus connecting the two lateral masses of gray matter across the midline, it is known as the *central canal*. The gray isthmus itself is called the *gray commissure*. Sometimes the parts in front and behind the central canal are referred to as the anterior and posterior gray commissures, respectively (fig 66 1).

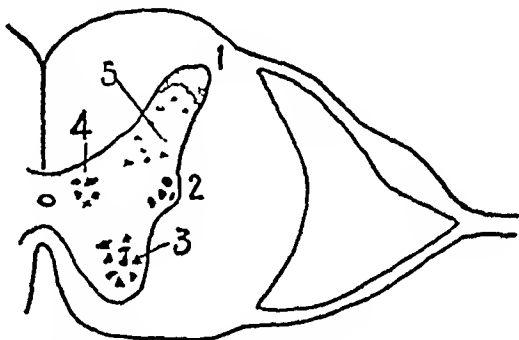


FIG 66 1 Main cell groups in the thoracic region of the spinal cord 1, substantia gelatinosa of Rolando, 2, intermedio lateral cell column (sympathetic), 3, cells (motoneurons) of anterior horn, 4, Clarke's column (dorsal nucleus) of posterior horn, 5, chief sensory nucleus

The *white matter*, which completely surrounds the gray matter is composed of bundles of fibers both myelinated and unmyelinated, the former predominating. A deep cleft on the ventral aspect of the cord (the *anterior median fissure*) and a septum upon the dorsal aspect (*posterior median septum*) together, incompletely divide the white matter into two lateral halves. The bands of white matter lying in front of the gray commissure is called the *anterior (ventral) white commissure*, a few strands of white matter lying behind or within the posterior part of the gray commissure is sometimes called the *posterior (dorsal) white commissure*. Each half is further marked out, by the fibers of the ventral and dorsal nerve roots, into three white columns or funiculi, ventral (anterior), lateral and dorsal (posterior). The *anterior (ventral) funiculus*

lies between the anterior median fissure on the one hand, and the ventral gray column (anterior horn) and the fibers of the ventral roots on the other. Bounded in front and medially by the last two structures, and postero-medially by the dorsal gray column and the fibers of the dorsal roots lies the *lateral funiculus*. The *dorsal funiculus* is situated between the dorsal gray column (posterior horn) and the dorsal root fibers which form its antero-lateral boundary, and the posterior median septum.

THE SPINAL NERVE ROOTS

There are 31 pairs of spinal nerve roots, a pair from each spinal segment. The anterior and posterior roots join within the vertebral canal just beyond the ganglion to form a corresponding number of spinal nerves.

The *anterior roots* of the spinal nerves are composed entirely of efferent fibers (p 911).¹ These are, (a) coarse, heavily myelinated, the axons of cells of the anterior horns, which are conveyed in the peripheral nerves to the skeletal muscles, and (b) fine, lightly myelinated fibers, in the thoracic and lumbar regions, these are preganglionic fibers of the sympathetic nervous system. The *posterior roots* are constituted of afferent fibers from the skin, muscles and viscera. They are the central processes of the large unipolar cells of the spinal ganglia. There is no important evidence that they contain efferent fibers (See Bell-Magendie law, p 276 and also p 949). Two divisions of the posterior root, a medial and a lateral, are distinguished. The lateral division is largely composed of small unmyelinated fibers, these, after entering the cord

form the small triangular area at the tip of the posterior horn known as Lissauer's tract (p 988). The fibers of the medial division, which are for the most part heavily myelinated, enter the posterior columns of the cord (see fig 66 5).

THE SEGMENTAL DISTRIBUTION OF THE SPINAL NERVES

In the young mammalian embryo and in certain adult lower forms, e.g., fishes, the body is demarcated into a regular series of transverse segments or *metameres*. The muscles (*myotomes*), skin (*dermatomes*) and viscera of each of these primitive blocks eventually receive innervation from the nerve roots of a corresponding spinal segment. The anterior root of each spinal nerve supplies motor fibers to the respective myotome, and autonomic fibers to the viscera and skin, the posterior root supplies sensory fibers to the corresponding dermatome as well as to the muscles and viscera. As a result of the outgrowth of the embryonic limbs the orderly arrangement of the metameres from before backwards becomes altered. In the adult mammal, the primitive metameric disposition is observed only in the trunk. The fibers of the spinal nerves supplying the limbs have joined to form the brachial and lumbosacral plexuses and, after intermingling freely, issue again as the peripheral nerves. The latter, in consequence, are composed of fibers derived from two or more spinal segments, and fibers from a given segment pass into several peripheral nerves. The muscles supplied by a given spinal segment do not necessarily lie in close proximity to one another (the coracobrachialis, for example, is innervated by the same segments as those which supply the muscles of the thumb) and a single muscle may derive its nerve supply from more than one spinal segment (see page 1006). As development proceeds and the limbs grow out from the trunk, the dermatomes become arranged in a series of narrow areas lying for the most part in the long axis of the limb (fig 66 2). The skin and muscles of the limbs also tend to move away from the visceral structures with which they were originally associated and, in the adult, structures innervated by a common spinal segment may be widely separated. Thus, the diaphragm is innervated (through the phrenic) from the 3, 4 and 5 cervical segments which also supply skin and muscle in the region of the neck and shoulder. The heart receives sensory and autonomic fibers from

¹Though doubt has been expressed from time to time concerning the purely efferent nature of the fibers composing the anterior roots, no substantial evidence has been ever brought forward to show that this is not so. It is true that stimulation of the central end of an anterior root sometimes gives rise to pain, but this is due to the presence of recurrent fibers from the posterior roots and in no way invalidates the Bell-Magendie law. Neurons have been found in the anterior roots resembling those of the posterior root ganglia, but there is no evidence that they are sensory in function. Some authors have been led to believe that the anterior roots transmit sensory impulses because pain is not always relieved by section of the posterior roots. There are three possible reasons for the failure of this operation. In the first place, the pain may arise within the central nervous system itself, a filament of a posterior root may have escaped division, or innervation from adjacent spinal segments may overlap to an unusual extent, pain would then continue to be registered from the area innervated by the severed roots.

the upper thoracic segments, these segments also supply sensory fibers to the skin over the inner aspect of the arm and hand and upper part of the thorax. The distribution of the dermatomes and cutaneous nerves in the human subject are shown in figure 66.3

Several methods have been employed by different investigators in mapping out the dermatomes in animals and in the human subject. The *anatomical* method is laborious and consists in tracing the fibers of a spinal root to their terminations in the skin. A *physiological* method ("isolation" or "sensory remainder" method) was employed by

vascular reaction (vasodilatation) as the means of demarcating the dermatomes

THE TRACTS OF THE CORD

The fiber tracts of the cord are divisible into two main groups (a) *long tracts* (projection tracts) which connect the cord with other parts of the central nervous system. Some of these (ascending tracts) carry impulses to higher centers, other tracts (descending) conduct in the reverse direction—from higher centers to spinal neurons. (b) *short tracts* (intersegmental or association tracts, ground bundles) which begin and end within the

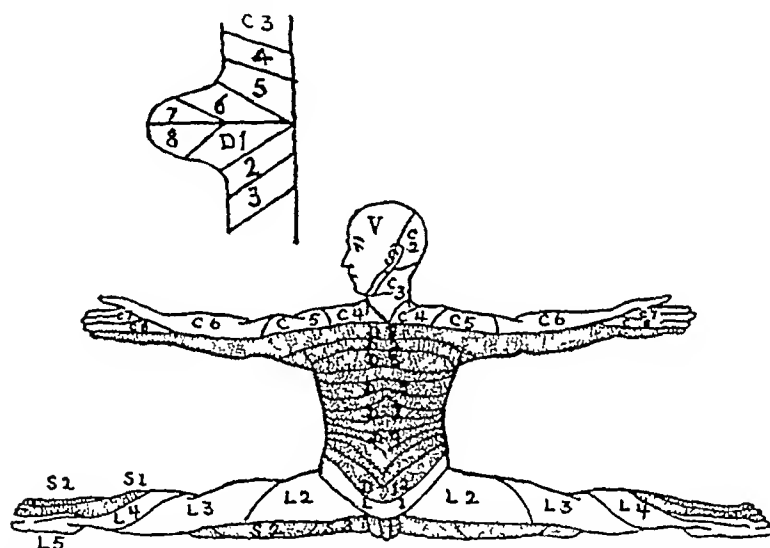


FIG 66.2 Upper Showing the drawing out of the metameres of the embryo with the development of the upper limb-bud Lower Showing the segmental arrangement of the dermatomes (From Strong and Elwyn, after Luciani)

Sherrington in cats and monkeys. The area of skin supplied by a given segment was demarcated by dividing the sensory roots above and below it. The sensitive area of skin bounded above and below by an anesthetic zone indicated the area of distribution of the undivided roots. Owing to the overlap of fibers from adjacent segments, it is not possible to produce an anesthetic area by the division of the sensory roots of a single segment (fig 66.4). Head mapped out the segmental distribution of the cutaneous nerves in the human subject from studies of cases of herpes zoster, a condition due to a lesion of the ganglion cells of the posterior roots. Within more recent years Foerster, using Sherrington's method, has mapped out the dermatomes in man. He also stimulated the posterior roots at operation and used the resulting

cord and connect different spinal segments. The fiber tracts making up the substance of the respective funiculi are listed in table 92 and shown diagrammatically in figure 66.5

ASCENDING TRACTS OF THE CORD

(1) **THE DORSOLATERAL (OR POSTERO) FASCICULUS** (TRACT OF LISSAUER) This is seen in cross section as a small area lying between the tip of the posterior horn and the periphery of the cord (fig 66.5). It is composed of fibers derived from the lateral division of the posterior nerve roots. These fibers upon entering the cord connect immediately, or after a very short upward or downward course, with cells occupying the tip of the posterior horn, i.e., in the substantia gelatinosa of Rolando (p 986). The fibers of this tract are mostly of small

diameter and for the most part unmyelinated, and from both pathological and experimental evidence there is little doubt but that they constitute the primary neurons in the pathway for pain and crude

columns of the cord The fibers of the *lateral* tract arise, as mentioned above, from cells in the substantia gelatinosa of Rolando, of the opposite side of the cord Those of the *ventral* tract are the

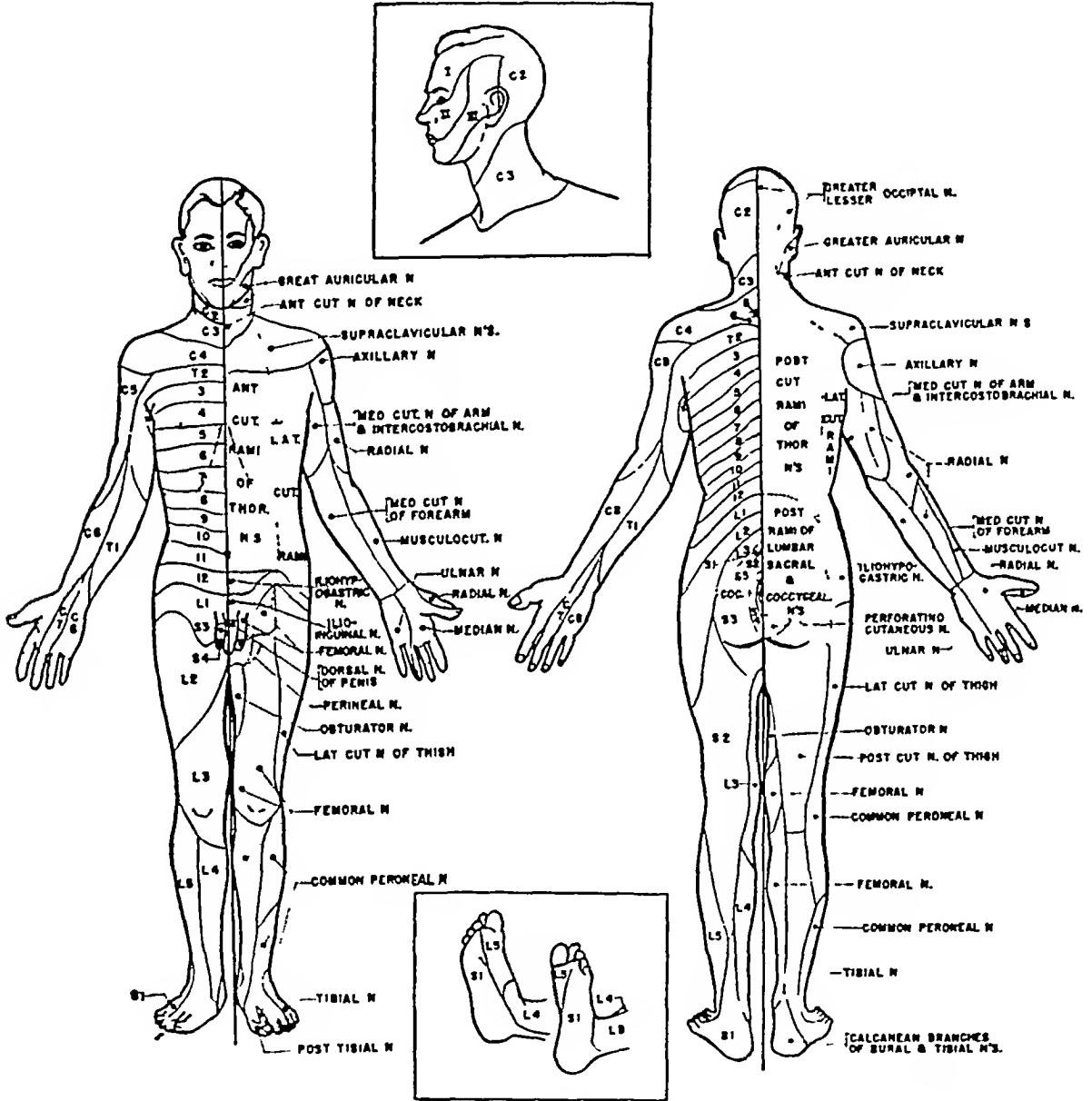


FIG 66.3 Showing distribution of cutaneous nerves (right halves of figures) and the dermatomes (left halves) (After Wolff)

thermal sensations The axons of the secondary neurons (axons of the cells in the substantia gelatinosa of Rolando just mentioned) go to form the lateral spinothalamic tract of the opposite side, to be immediately described

(2) THE LATERAL (DORSAL) AND VENTRAL (ANTERIOR) SPINOTHALAMIC TRACTS These tracts occupy, respectively, the lateral and ventral

axons of cells in the chief sensory nucleus of the opposite dorsal horn of gray matter

In order to reach the ventral spinothalamic tract of the opposite side, the fibers arising from the chief sensory nucleus ascend in the posterior columns for two or three spinal segments, and then cross in the anterior white commissure This tract is a crossed pathway for the sensations of *touch*,

and probably also for *tactile localization*. These sensations have, therefore, a double path to consciousness for they are also conveyed uncrossed in the posterior columns of the cord.

The fibers going to form the lateral spinothalamic tract (secondary neurons from cells of substantia gelatinosa of Rolando of the opposite side) ascend for a short way (within a single segment), and then cross in the anterior white commissure. This tract transmits impulses aroused by all forms of thermal and painful stimuli.

The two tracts—anterior and lateral—come together in the medulla to constitute the spinal lemniscus which joins the medial lemniscus in the upper part of the medulla to enter the thalamus (ch 67).

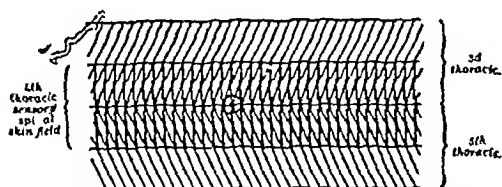


FIG 664. Diagram showing the overlapping of the skin area innervated through the fourth thoracic spinal nerve root by those supplied by the third and fifth. The small circle indicates the position of the nipple (after Sherrington)

(3) THE SPINOTECTAL TRACT is placed in the lateral column ventral to the lateral spinothalamic tract. Its fibers arise from cells in the posterior horn of the opposite side and terminate in the superior colliculus. It subserves spinovisual reflexes.

(4) SPINO-OLIVARY TRACT. The fibers of this tract are the axons of cells in the dorsal horn of gray matter; they run with olivo spinal fibers and end in the inferior olivary nucleus of the opposite side. They probably transmit proprioceptive impulses which are relayed from the olivary nucleus to the cerebellum.

(5) SPINOCEREBELLAR TRACTS. (a) The dorsal (posterior) *spinocerebellar tract* (direct cerebellar or tract of Flechsig) is situated in the posterior part of the lateral funiculus to the outer side of the lateral corticospinal tract. It is composed of the axons of the cells of Clarke's column (p 986) of the same side and, possibly, to some extent of the opposite side. The direct cerebellar tract reaches the cerebellum via the inferior cerebellar peduncle; its fibers end mainly in the cortex of the anterior and posterior lobes of the cerebellum (ch 70). (b)

The *ventral (anterior or indirect) cerebellar tract* arises from cells in the base of the dorsal horn, mainly of the same side, but also of the opposite side of the cord. It ascends in front of the dorsal spinocerebellar tract and is continued upwards through the brain stem as far as the mid-brain where it arches backwards and downwards (arciform fibers) to reach the cerebellum via its superior peduncle. Its constituent fibers end in the cortex of the anterior and posterior cerebellar lobes. The spinocerebellar tracts carry impulses arising in the proprioceptors of the muscles, tendons and joints. The information thus conveyed to the cerebellum is essential for the latter's function in adjusting the tone of the skeletal muscles and synergizing their movements (ch 70).

(6) THE FASCICULI GRACILIS (TRACT OF GOLL) AND CUNEATUS (TRACT OF BURDACH) occupy the dorsal (posterior) funiculus of the cord. In the upper part of the cord the former tract lies on the medial side of the latter.

Both tracts are composed of heavily myelinated fibers which are the continuation upwards of the fibers of the medial division of the posterior nerve roots of the same side. The fasciculus gracilis commences in the lowest level of the spinal cord. After their entrance into the cord, its fibers divide into long ascending and short descending branches. A few of the former fibers and all the latter, after a short course, enter the gray matter. The long divisions become displaced medially as they ascend, with the result that fibers arising at lower levels (e.g., sacral region) come to lie nearer the midline than those entering the cord higher up. The fasciculus cuneatus first appears in the mid-thoracic region. Therefore, from the mid-thoracic region upwards, the more medially placed tract (fasciculus gracilis) is derived from the lower thoracic, and the lumbar and sacral nerve roots, in this order, from the lateral to the medial border of the tract. The ascending fibers of both tracts, except those mentioned above as constituting the ventral spinothalamic tracts, pass uncrossed to the medulla, ending, respectively, in the nucleus gracilis and the nucleus cuneatus. From these nuclei the axons of secondary neurons emerge and, passing medially as the *internal arcuate fibers*, decussate with those of the opposite side (sensory decussation). They ascend through the brain stem as the *medial lemniscus* (p 993) to terminate in the posteroventral part of the lateral nucleus of the thalamus. Other fibers (*external arcuate*) from the

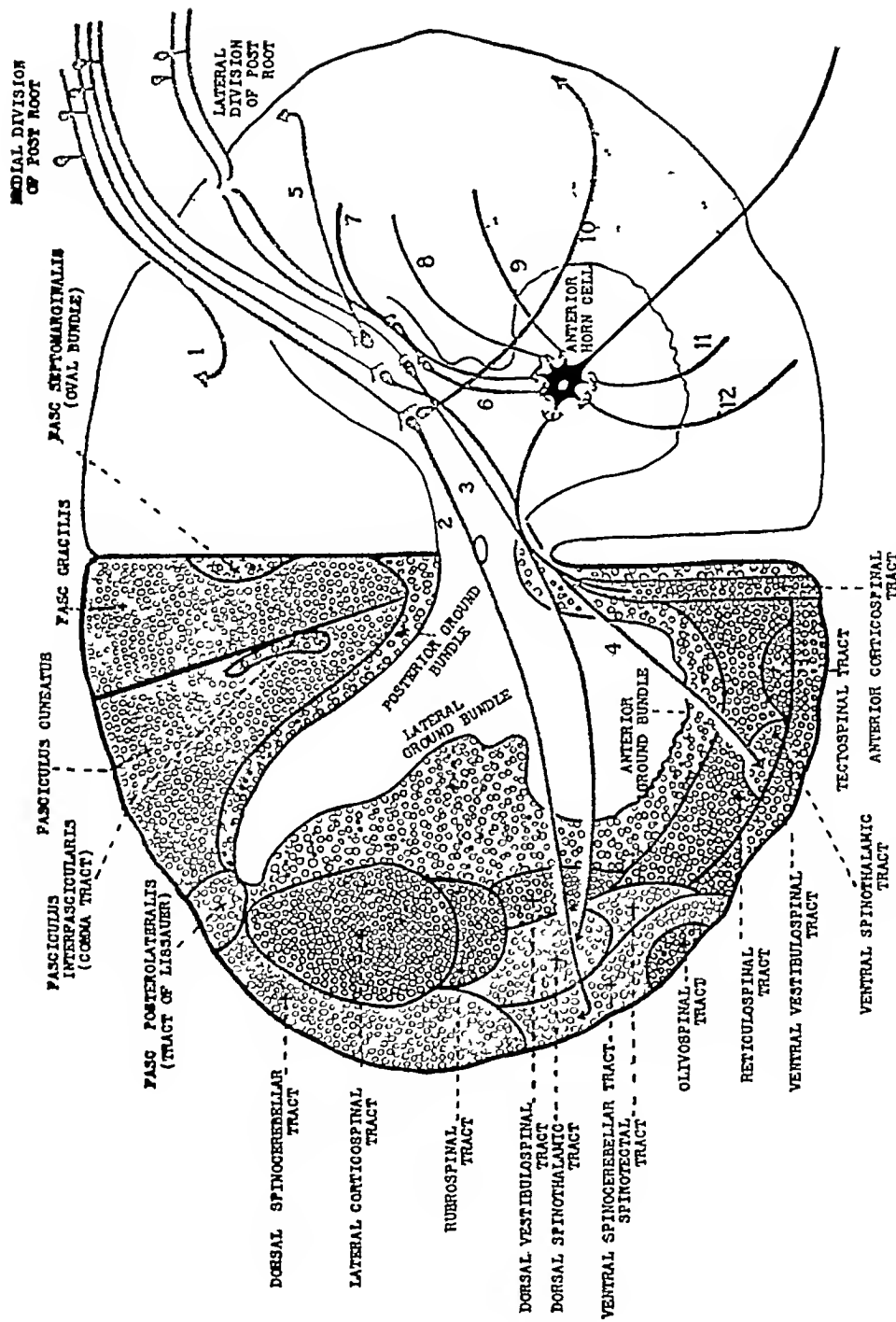


Fig 66.5 Diagram to show tracts of the cord. Ascending fibers shown in blue, descending in red, spinal motoneuron in black 1, represents fibers ascending in posterior columns (mediating sensations of touch, spatial discrimination, and of position and movement), 2 and 10, represent fibers entering the contralateral and homolateral ventral spinocerebellar tracts, respectively, 3 represents fibers entering dorsal spinotthalamic tract of the opposite side (mediating pain and thermal sensations), 4, represents fibers entering ventral spinotthalamic tract of the opposite side (mediating touch and tactile localization), 5, fiber entering dorsal spinocerebellar tract, 6, internuncial neuron connecting a posterior root fiber with anterior horn cell (reflex arc), 7 and 8, crossed corticospinal fibers connecting, respectively, through an internuncial neuron and directly, with anterior horn cell, fiber of anterior corticospinal tract (not numbered) also shown, 9, rubrospinal fiber, 11, reticulospinal fiber, 12, vestibulospinal fiber

nuclei gracilis and cuneatus are relayed to the cerebellum (p 993) nonsensory impulses brought to them by the corresponding spinal fasciculi

they are unmyelinated Within the cord they constitute Lissauer's tract Immediately, or after a very short upward or downward course, they enter

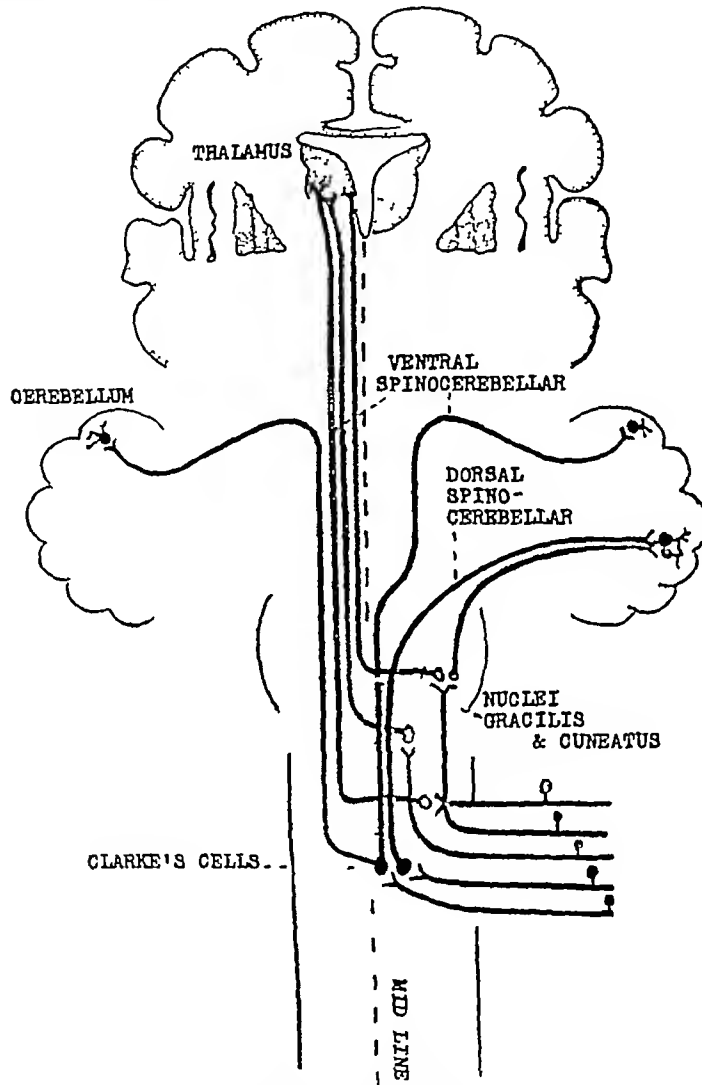


FIG 66 6 Diagram showing the course of afferent impulses after their entrance into the cord *Red*, pathway for pain and thermal sensations (ascend in dorsal spinothalamic tract of opposite side) *Blue*, pathway for touch, sense of position and movement, and spatial discrimination (ascend in posterior columns), external arcuate fibers from posterior column nuclei to cerebellum also shown *Green*, pathway for touch, and tactile localization (after travelling for variable distances in posterior columns, fibers cross to ventral spinothalamic tract of opposite side) *Black*, ventral and dorsal spinocerebellar tracts, some fibers of latter (not shown) also cross to opposite side

SUMMARY OF THE PRINCIPAL PATHWAYS ASCENDING THROUGH THE CORD

A SENSORY (conveying impulses to consciousness) (1) The fibers mediating thermal and painful sensations enter the cord in the lateral divisions of the posterior nerve roots According to Ranson

the dorsal horn and connect with nerve cells in the substantia gelatinosa of Rolando The axons of the latter secondary neurons cross to the opposite side in the white commissures and ascend as the lateral spinothalamic tract This tract occupies a position in the brain stem lateral to the medial

lemniscus and ends in the lateral nucleus of the thalamus (cf figs 66.5 and 66.6)

(2) The fibers conveying all other sensory impulses, e.g., from the muscles and joints (sense of movement and position), and those mediating the senses of vibration, touch, pressure, tactile localization and spatial discrimination (compass test), enter the cord in the medial division of the posterior roots and ascend in the fasciculus gracilis and fasciculus cuneatus. The fibers mediating all these sensations, except *some* of those concerned with touch and probably with tactile localization, pass without crossing to the nuclei in the medulla where they connect with secondary neurons. The axons of the latter constitute the *internal arcuate* fibers; they decussate with those of the opposite side and ascend to the thalamus as the medial lemniscus (or fillet). Tertiary neurons convey the impulses to the cerebral cortex.

In a transverse spinal lesion interrupting solely the posterior fasciculi, even if bilateral, the sensation of touch is retained, since the fibers which have entered the gray matter and crossed below the level of the lesion to ascend in the ventral spinothalamic tract and thus escape injury (see also hemi-section of the cord).

B NON-SENSORY impulses from the muscles, tendons and joints as well as non-sensory impulses aroused by a tactile stimulus, are conveyed into the cord by fibers composing the medial divisions of the posterior roots. These fibers connect immediately with the cells of Clarke's column and are continued upwards in the dorsal (direct) and ventral (indirect) cerebellar tracts of the same side, mainly, but also of the opposite side. Some of these impulses, viz., those which enter the cervical cord, are conveyed by fibers of the cuneate fasciculus to the cuneate nucleus, to be relayed by secondary neurons (*external arcuate fibers*) via the inferior peduncles to the cerebellum (see fig 66.6).

SENSORY PATHS IN THE BRAIN STEM (MEDULLA, PONS AND MID-BRAIN)

The *medial lemniscus* (or fillet). This is constituted of fibers arising in the nuclei gracilis and cuneatus. The fibers leave the ventral aspects of these nuclei and arch forward and medially (as *internal arcuate fibers*) to the midline where they cross with corresponding fibers of the opposite side (*sensory decussation*). They then turn upwards as a compact bundle known as the medial lemniscus or fillet. This ascends through the medulla and

pons dorsal to the pyramidal tracts, and through the tegmentum of the mid-brain.

The *spinal lemniscus* (or fillet) is formed by the fusion of the anterior and lateral spinothalamic tracts and is therefore a crossed path for impulses aroused by light touch, pressure, pain, heat and cold. In the upper part of the medulla the spinal lemniscus joins the medial lemniscus on its outer side, and is joined in the pons by the trigeminal lemniscus.

The *trigeminal lemniscus* (or fillet) conveys impulses from the area of distribution of the trigeminal nerve of the opposite side (p. 997).

The fibers of the medial, spinal and trigeminal lemnisci terminate in the thalamus, from where tertiary neurons pass to the cerebral cortex (see fig 66.7).

The *lateral lemniscus* constitutes the pathway for auditory impulses from the cochlear nuclei to the inferior colliculus and medial geniculate body (ch. 77).

The sensory pathways are constituted of three neurons. For example, the pathway traversed by the sensory impulses for touch or kinesthetic sense consists of a primary neuron whose cell body lies in the posterior root ganglion, a secondary neuron originating in a posterior column nucleus (gracilis or cuneatus) and a tertiary neuron arising in the thalamus which convey impulses to the somesthetic area of the cerebral cortex (ch. 68). The cells of the Gasserian ganglia are the primary neurons of the trigeminal pathway, secondary neurons lie in the sensory nucleus and in the spinal nucleus of the nerve (p. 998), tertiary neuron fibers ascend to the cortex from the thalamus. In the upper part of the pons and in the mid-brain these several sensory pathways become fused together into a compact bundle. But a lesion in the lower part of the brain stem may involve one of the sensory pathways exclusively of the others. Thus an injury localized to the outer part of the lower pons or of the medulla may by injuring the spinal lemniscus cause loss of sensation to pain, heat and cold over the *opposite half* of the body leaving muscle sense and tactile discrimination intact. Sensory loss of this nature accompanied by cerebellar symptoms occurs as a result of the occlusion (as by thrombosis or embolism) of the posterior inferior cerebellar artery. Usually also, as a result of the involvement of the spinal tract of the trigeminal nerve (p. 997) the face on the *same side* as the occluded vessel shows the dissociated sensory loss. A lesion more

centrally placed may, by implicating the medial lemniscus alone, cause the converse type of dissociated sensory defect, namely, loss of the sense of position of the limbs and of spatial discrimination with retention of sensibility to pain, heat and cold

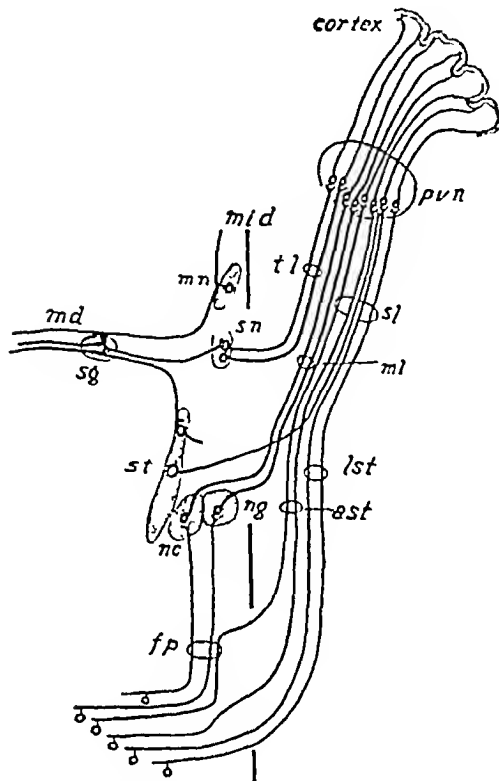


FIG 667 Diagram of the trigeminal medial and spinal lemnisci *mn*, mesencephalic nucleus of trigeminal nerve, *sn*, sensory nuc of trigeminal, *sl* spinal tract and nucleus of trigeminal, *md*, fiber of mesencephalic root passing into mandibular nerve, *sg* semilunar ganglion, *nc* nuc. cuneatus, *fp*, posterior columns of the cord, *ng* nuc. gracilis, *tl*, trigeminal lemniscus, *prn*, posterior ventral nuc of the thalamus, *st*, spinal lemniscus, *ml*, medial lemniscus, *lst*, lateral spinothalamic tract, *ast*, anterior spinothalamic tract, *mid*, mid line. The trigeminal is on the medial side of the medial lemniscus, only in the medulla, it comes to lie on the outer side in the pons, and in close relation to the spinal lemniscus

In lesions at higher levels in the brain stem all forms of sensation are likely to be involved more or less equally

DESCENDING TRACTS OF THE CORD

(1) THE CORTICOSPINAL (CEREBROSPINAL) OR PYRAMIDAL TRACTS From their origins in the cerebrum and brain stem on each side (see also

p 1002) to the lower border of the medulla. Here each tract divides into two bundles of unequal size, the larger of which crosses with the corresponding bundle of the opposite side and descends in the posterior part of the lateral funiculus of the cord as the *lateral corticospinal tract* (*crossed pyramidal tract*). The remaining fibers, uncrossed, descend in the anterior funiculus as the *anterior corticospinal tract* (*direct pyramidal tract*). They constitute only from 10 to 20 per cent of the total number of corticospinal fibers. The direct pyramidal tract is well marked in the cervical region but its fibers dwindle gradually in number, in successive segments being scarce or absent below the mid-thoracic region, though a few may be found as low as the sacral region. A proportion of the fibers of this tract also cross eventually to the opposite side of the cord, passing at different levels through the anterior white commissure (cf figs 66.5 and 66.8). Besides the well recognized anterior (direct) corticospinal tract there is, as shown by Schuster and lately by Fulton and Sheehan another tract of uncrossed pyramidal fibers to be found in the lateral column of the cord. Through these two tracts (anterior and lateral) of uncrossed fibers the skeletal muscles of each side of the body receive impulses from both motor areas of the cortex.

All corticospinal fibers whether crossed or uncrossed connect with the large motor cells of the ventral gray columns (anterior horn cells). The connections are of two types, (a) direct synapses (20 to 30 per cent) with the motor neurons, (b) in direct connections with these cells through an internuncial neuron whose cell body is also in communication with a posterior root fiber on each side. It has been estimated that from 75 to 90 per cent of the corticospinal fibers terminate in the cervical (55 per cent) and thoracic (20 per cent) regions of the cord, and about 25 per cent in the lumbar and sacral regions.

Contrary to previous belief, the corticospinal or pyramidal tract is made up only to a very minor extent of the axons of the large pyramidal cells of Betz, see pp 1002 and 1030).

(2) THE VESTIBULOSPINAL TRACTS These are two in number, a ventral and a lateral. The *ventral vestibulospinal tract* descends in the forepart of the anterior funiculus and lateral to the anterior portion of the direct pyramidal tract. Its fibers arise from the lateral vestibular nucleus (Dietz's) chiefly of the same side (though some are crossed). They may be traced as far as the sacral region. The fibers terminate by forming synapses with the

motor neurons either directly or through an internuncial neuron. The vestibulospinal tract forms a section of the pathway from the labyrinth and the cerebellum to the skeletal muscles.

(3) **THE TECTOSPINAL TRACTS** of each spinal half arise from the contralateral superior and inferior colliculi and descend at first as a single compact bundle through the reticular formation of the brain stem. Upon reaching the cord the fibers are segregated into two fasciculi, a ventral and a dorsal, which descend in the ventral and lateral white columns respectively. The fibers synapse with the spinal motoneurons, either

ascend through the reticular formation of the pons and medulla to enter the lateral funiculus of the cord ventral to the lateral corticospinal tract. In man, they cannot be traced below the mid-thoracic segments of the cord. Here the rubrospinal fibers connect either directly or through internuncial neurons with the motor neurons, to which they relay impulses from the cerebellum and striate body.

(5) **THE OLIVOSPINAL (BULBOSPINAL) TRACT OF HELWEG.** Its fibers arise from cells in the neighborhood of the inferior olivary nucleus and descend in the ventral and lateral part of the lateral funicu-

TABLE 93

FUNICULUS	ASCENDING TRACTS	DESCENDING TRACTS	INTERSEGMENTAL TRACTS (GROUND BUNDLES)
Ventral (anterior)	Ventral (anterior) spinothalamic	Ventral (anterior) corticospinal (direct pyramidal) Vestibulospinal Ventral (anterior) tectospinal Reticulospinal	Ventral (anterior) intersegmental fasciculus
Lateral	Lateral spinothalamic Dorsal (posterior) spinocerebellar (Flechsig) Ventral (anterior) spinocerebellar Spinotectal Dorsolateral (posterolateral) fasciculus (Lissauer)	Lateral corticospinal (crossed pyramidal) Rubrospinal Olivospinal (Helweg) Dorsal tectospinal	Lateral intersegmental fasciculus
Dorsal (posterior)	Fasciculus gracilis Fasciculus cuneatus		Septomarginal fasciculus Dorsal (posterior) intersegmental fasciculus

directly or through internuncial neurons, especially in the cervical region, wherein are situated the centers for the neck muscles. Impulses from the retina are received by the superior colliculus which through such impulses, and its spinal connections, serves as a center for the integration of visual impressions with body movement, especially of the head (*visuospatial reflexes*). This center is probably of less importance in man, with his highly developed visual cortex, than in animals. Auditory reflexes may also be mediated by this tract and the inferior colliculus.

(4) **THE RUBROSPINAL TRACT** arises from the large cells (nucleus magnocellularis) in the posterior part of the red nucleus (p. 1017). The fibers immediately upon leaving the red nucleus cross to the opposite side (*Forel's decussation*) and de-

scend no farther than the cervical region. The fibers synapse with motor neurons. The functions of this tract are unknown. It and the thalamo-olivary tract constitute a possible path whereby impulses from the thalamus may reach the spinal centers.

(6) **THE RETICULOSPINAL TRACTS** arise from cells scattered through the reticular formation mainly of the upper part of the pons but also of the midbrain and medulla. They are two in number, one in which crossed fibers predominate descends in the anterior funiculus, the other composed mainly of uncrossed fibers, in the lateral funiculus. They connect with the motor neurons in the anterior horns and probably with cells in the lateral horn. These tracts relay impulses from the striate body and cerebellar cortex and from the

small-celled nucleus (n parvocellularis) of the red nucleus (via the *rubroreticular tract*)

INTERSEGMENTAL TRACTS OF THE CORD AND THE MEDIAL LONGITUDINAL FASCICULUS

There is in each funiculus of the cord lying close to the gray matter an intersegmental fasciculus or ground bundle. These, as already men-

tioned, serve to link up spinal segments of different levels. Their fibers arise from cells in the gray matter and after an ascending or descending course of variable length end around cells of the same or of the opposite side at a higher or a lower level. A proportion of the fibers constituting the intersegmental fasciculus of the lateral column of the cord (*lateral intersegmental fasciculus* or *lateral ground bundle*) are continued upwards into the medial longitudinal fasciculus (or bundle) of the

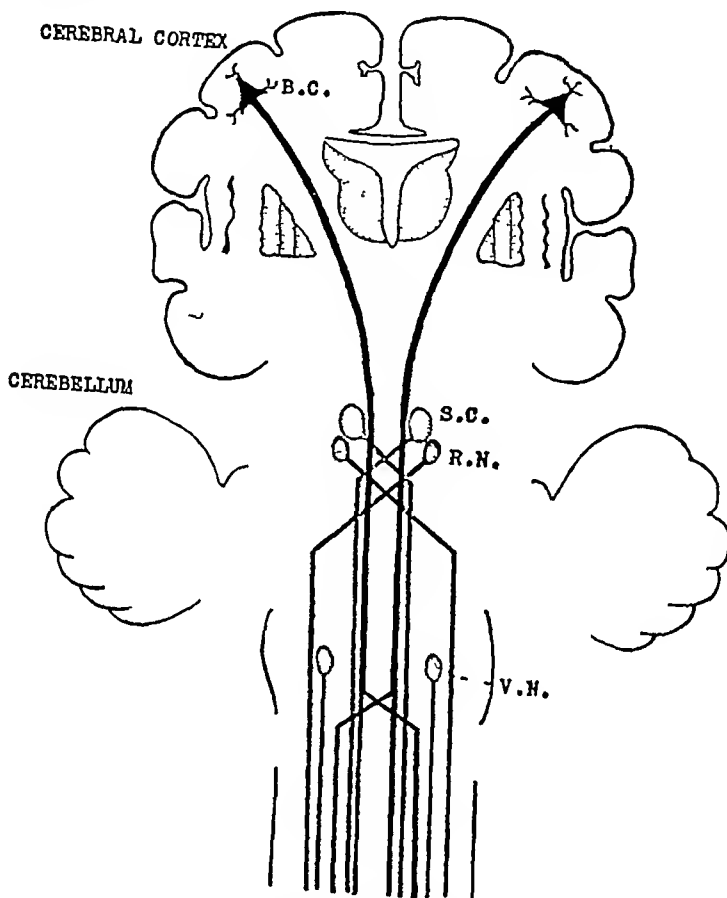


FIG 66.8 Diagram of the descending tracts of the cord showing Pyramidal tracts (black), rubrospinal tracts (red), tectospinal tracts (blue) and vestibulospinal tracts (green) B.C., Betz cell of motor cortex, R.N., red nucleus, S.C., superior colliculus, V.N., vestibular nucleus

tioned, serve to link up spinal segments of different levels. Their fibers arise from cells in the gray matter and after an ascending or descending course of variable length end around cells of the same or of the opposite side at a higher or a lower level. A proportion of the fibers constituting the intersegmental fasciculus of the lateral column of the cord (*lateral intersegmental fasciculus* or *lateral ground bundle*) are continued upwards into the medial longitudinal fasciculus (or bundle) of the

The posterior funiculus contains (a) the septomarginal fasciculus and (b) the posterior intersegmental fasciculus (or posterior ground bundle).

The *septo-marginal fasciculus* is composed of (a) intersegmental fibers which arise from cells of the posterior horn and synapse with corresponding cells at lower levels and (b) descending fibers of the medial divisions of the posterior nerve roots. Septomarginal fasciculus is a term applicable to these fibers only in the lumbar region where they abut on each side against the posterior median septum. They are also known here

as the *oval area* or *bundle of Flechsig*. These fibers form areas of different sizes and shapes on cross section at various spinal levels. In the cervical and upper thoracic regions the fibers appear on section as a crescentic area at about the center of the posterior funiculus. Here they are spoken of as the *comma tract of Schultz*, or *tractus interfascicularis*. In the lower thoracic region they form a narrow zone bounding the posterior aspect of the cord and are known as the *dorsal peripheral strand*. In the sacral region this tract lies against the posterior part of the median septum and is called the triangular area of Philippe-Gombault. The descending fibers of the branches of the medial division of the posterior roots as they descend also, like the long ascending branches, become displaced toward the midline.

The *posterior intersegmental fasciculus* is seen in cross section as a small area lying behind the posterior gray commissure. Its fibers connect the posterior horn cells of different segments (see fig 66 5).

THE LONGITUDINAL FASCICULUS (OR BUNDLE)

The medial (posterior) longitudinal fasciculus (or bundle) is a tract of great physiological importance and is present in all vertebrates. It is composed of fibers which connect the 3rd, 4th and 6th cranial nerves with one another and with the vestibular nuclei, and makes connections with the motoneurons of the cervical segments of the spinal cord (fig 65 15). The tract lies near the mid-line and extends from the cervical segments of the cord, where it is continuous with the anterior intersegmental fasciculus (anterior ground bundle), to the floor of the 3rd ventricle. In the medulla it lies immediately subjacent to the floor of the 4th ventricle, in the pons it courses through the *formatio reticularis*, and in the mid-brain, lies in relation to the gray matter of the floor of the Sylvian (cerebral) aqueduct. It also received fibers from the lateral lemniscus (auditory path, p 1189) and from the superior colliculus, through which it is in communication with the optic pathway (p 1170). The essential function of the medial longitudinal bundle is the coordination of reflex movements of the ocular and neck muscles in response to labyrinthine, auditory and visual stimuli.

THE CENTRAL CONNECTIONS OF THE TRIGEMINAL, FACIAL, GLOSSOPHARYNGEAL, VAGUS, ACCESSORY AND HYPOGLOSSAL NERVES

THE TRIGEMINAL PATHWAY

The trigeminal nerve appears on the ventral surface of the pons a little to one side of the midline, and somewhat below the superior pontine

border. The nerve has three roots (a) a large *sensory*, (b) a small *motor*, and (c) a *mesencephalic* (fig 66 9).

The fibers of the *sensory root* convey impulses from the anterior part of the scalp from the skin of the forehead and face, with the exception of an area over the angle and lower border of the mandible. It also supplies the mucous membrane of the mouth (anterior two-thirds of tongue) and nose, the cornea and conjunctivae and the dura mater

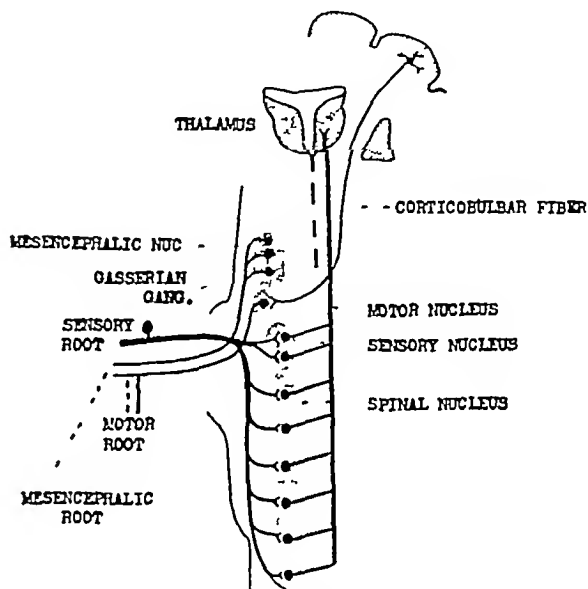


FIG 66 9 Diagram of the central connections of the trigeminal nerve

This root arises from the cells of the Gasserian (semilunar) ganglion, which corresponds to a spinal ganglion. The processes of the ganglion cells divide into peripheral and central branches. The former enter into the composition of the peripheral divisions (ophthalmic, maxillary and mandibular) of the nerve, the central branches (which constitute the sensory root of the nerve) enter the pons in close association with the fibers of the motor root and divide into an ascending and a descending group in a manner homologous with a spinal root fiber. The *ascending fibers* convey impulses of light touch, tactile discrimination and localization, and of the sense of position and passive movement. They end in a nucleus (*main* or *upper sensory nucleus of the trigeminal*) situated in the pons, deep and lateral to the motor nucleus. The axons of cells in this nucleus ascend with those from the spinal nucleus of the trigeminal (see below) to form the *trigeminal lemniscus*. The *descending fibers* constitute the *bulbosplinal tract of the trigeminal*. They

subserve sensations of pain and temperature over the entire trigeminal area. They descend through the pons and medulla and may be traced as far as the 2nd cervical segment. This tract dwindles gradually in its descent, its fibers terminating around cells in the substantia gelatinosa of Rolando which in this situation is referred to as *spinal nucleus of the trigeminal*. The fibers of the mandibular, maxillary and ophthalmic divisions of the trigeminal end in the nucleus in this order from above downwards, that is, in an order the inverse of that in which they are distributed to the skin of the face. From the spinal nucleus the axons of secondary neurons emerge which cross the mid-line, ascending, they join fibers arising from the upper sensory nucleus to form the *trigeminal lemniscus* (or *fillet*). The trigeminal lemniscus lies in close association with the medial and the spinal lemniscus (spinothalamic tracts) (p. 989), its fibers terminate in the posteroventral nucleus of the thalamus.

A number of reflexes are mediated through the afferent fibers of the trigeminal nerve, e.g., corneal, winking, sneezing and oculocardiac.

The main sensory nucleus of the trigeminal corresponds to the nuclei of the posterior columns of the cord (nuclei gracilis and cuneatus) and like the latter sends fibers to the cerebellum through the inferior cerebellar peduncle. The spinal tract of the trigeminal may be looked upon as corresponding to the tract of Lissauer (p. 988), and the spinal nucleus as an extension upwards of the substantia gelatinosa of Rolando whose cells, as we have seen, give rise to the lateral spinothalamic tract.

The *motor root* fibers arise from a nucleus in the upper part of the pons underlying the lateral part of the floor of the 4th ventricle. The nucleus receives fibers from the corticobulbar tract of the opposite side and also probably from the same side. The motor root after its emergence from the brain travels peripherally with the sensory root and, passing deep to the Gasserian (semilunar) ganglion, joins the mandibular (3rd) division of the trigeminal to supply the muscles of mastication (temporal, masseter and pterygoids).

The *mesencephalic root* consists of a small bundle of fibers which run in company with the fibers of the motor root. Entering the pons they ascend to the *mesencephalic nucleus* of the trigeminal, an elongated collection of nerve cells extending from the level of the motor nucleus to the upper region of the mid brain. It was thought at one time that the mesencephalic root and nucleus were motor in

function, but it is now generally admitted that they are composed of afferent neurons. The nucleus is looked upon as a group of cells homologous with the Gasserian (semilunar) ganglion and the posterior root ganglia of the spinal nerves, but which has migrated or been "drawn" into the brain at an early period of phylogenetic development. It is believed to receive proprioceptive impulses from the muscles of mastication and possibly from the ocular muscles.

LESIONS INVOLVING THE TRIGEMINAL PATHWAYS Either the peripheral portions or the central connections of the nerve may be the seat of disease. A lesion of the nerve peripheral to the ganglion is more likely to involve only one of its three divisions. Pain, loss of sensibility over the distribution of the division affected, may result or be more widespread if the ganglion cells are involved. Herpetic eruptions over the areas of distribution of all three divisions may result from involvement of the ganglion. Paralysis of the muscles of mastication may follow injury or disease of the mandibular division, or analgesia of the cornea with ulceration (neuropathic keratitis) may result from an affection of the ophthalmic division. Loss of sensation of taste (as well as of ordinary sensibility) over the anterior two thirds of the tongue on the corresponding side is a common accompaniment of degenerative changes affecting the mandibular division of the 5th nerve. The taste fibers to this part of the tongue are derived from the chorda tympani branch of the facial, but travel via the lingual branch of the mandibular division of the trigeminal. The loss of taste is due to pressure upon the chorda fibers by the degenerating lingual filaments, and is, as a rule, only temporary. Taste fibers are not constituents of the 5th nerve itself. Removal of the Gasserian ganglion, for instance, does not result in permanent loss of taste on the operated side. Degeneration of the nerve central to the ganglion may occur in tabes. Lesions (e.g., tumors, vascular changes) in the pons, medulla or upper cervical cord may injure the upper sensory nucleus, the motor nucleus, the trigeminal lemniscus (crossed) or the bulbospinal tract of the nerve. When the motor nucleus is involved weakness and wasting of the muscles of mastication result, implication of the ascending sensory fibers, or of the main sensory nucleus in the pons, is followed by loss of the sensation of light touch and the discriminative aspects of cutaneous sensibility over the same side of the face, but the retention of sensibility to pain, heat and cold. The neighboring

spinothalamic tract (crossed) may suffer coincidentally with the ascending sensory fibers, when thermanesthesia and analgesia over the trunk and limbs of the opposite side combined with loss of tactile sensation over the face of the same side will result (see also p 1009). Syringomyelia extending into the upper cervical region or into the bulb (syringobulbia) is likely to cause, as a result of pressure upon the spinal tract of the nerve, dissociated sensory loss over the face is the reverse of that caused by a lesion of the main sensory nucleus of the trigeminal—loss of the appreciation of pain and changes in temperature, with retention of tactile sensibility.

The trigeminal nerve, or one of its divisions, is sometimes the seat of a severe and intractable type of pain which recurs in paroxysms (*trigeminal neuralgia*) and may be accompanied by reflex spasms of the facial muscles (*tic douloureux*). The cause of the affection is unknown. In some instances it may be of thalamic origin. In treating the condition, injections of alcohol into the division of the nerve involved are sometimes employed. The injection is made into the nerve at the infraorbital foramen or foramen rotundum in the case of involvement of the maxillary division, and at the foramen ovale of the sphenoid or at the supra-orbital notch, respectively, in disease of the mandibular or ophthalmic division. Injections into the ganglion itself, or section of the sensory root before it enters the ganglion (i.e., intradurally) is more likely to be followed by permanent relief, injury to the motor root is avoided, and regeneration of the sensory fibers (which do not possess a neurilemma) proximal to the ganglion does not occur.

THE FACIAL NERVE

The facial nerve consists of a large *motor* and a small *sensory* portion. The two portions or roots appear at the lower border of the pons and enter the internal auditory meatus in company with the auditory nerve.

The *sensory* root of the facial, which is also known as the *nervus intermedius* of Wrisberg, contains not only afferent fibers but secretory and vasodilator (parasympathetic) fibers as well.

The *sensory fibers* arise from the cells of the genicular (geniculate) ganglion (fig 66 10). The peripheral processes of these cells are distributed through the chorda tympani branch of the facial nerve to the taste buds and mucous membrane of the anterior two-thirds of the tongue, and through

the nerve to the pterygoid canal (formed by the union of the greater superficial petrosal and the deep petrosal nerves), and the sphenopalatine ganglion to the lacrimal gland and the mucosa of the soft palate and posterior part of the nose. The chorda tympani branch joins the trunk of the lingual nerve through which it is conveyed to the floor of the mouth.² Deep sensibility (pressure, pain) from the facial muscles is also, according to Loyal Davis, conveyed by afferent fibers which are distributed with the motor fibers but pass centrally in the nervus intermedius. Others claim that all pain, deep and superficial, is carried by the trigeminal. The central processes of the ganglion cells end in the *sensory nucleus* situated in the upper part of the tractus solitarius. Fibers arise from the latter and ascend in the medial fillet of the opposite side to reach the thalamus (anterior nucleus). From here the pathway for taste impulses is continued by tertiary neurons to the lower part of the somesthetic area of the cerebral cortex.

The *parasympathetic fibers* arise from the *superior salivatory nucleus* (p 488) which lies in close relation to the motor nucleus of the facial. After leaving the brain stem in the sensory root the parasympathetic secretory and vasodilator fibers pass via the *great superficial petrosal nerve* and *nerve to pterygoid canal* (formed by the union of the great superficial and deep petrosal nerves) to the *sphenopalatine ganglion* from where they are relayed to the lacrimal gland, and to the vessels and glands of the palate and posterior part of the nose. The secretory and vasodilator fibers to the submaxillary and sublingual glands leave the facial with the taste fibers in the chorda tympani branch.

The *motor part* of the facial is conveyed through the facial canal of the temporal bone to the stylomastoid foramen. After its emergence from the latter it is distributed to muscles of the face, auricle, and forehead. The motor fibers arise from

² The nervus intermedius, the genicular ganglion, the chorda tympani and part of the great superficial petrosal nerve are sometimes grouped together under the name *glossopalatine nerve*. The origin and distribution of the secretory and sensory fibers of which this nerve is composed are closely similar to those of the glossopharyngeal, and it is considered by some as an aberrant part of the latter nerve.

A small proportion of taste fibers may take an alternative route, namely, via the chorda tympani to the otic ganglion and thence by way of the internal sphenoidal and great superficial petrosal nerves, genicular ganglion and nervus intermedius to the brain stem (Schwartz and Weddell).

a nucleus in the lower part of the pons and pass backwards to the lower end of the nucleus of the abducent nerve. Then ascending to the upper end of the latter they bend and sweep downwards and forward to where they leave the brain. As the facial fibers take this arched course to the point of their emergence from the brain stem, they, together with the abducent nucleus, form a

FACIAL PARALYSIS The effects of interruption of the facial pathway vary in certain important features according to the level at which the injury occurs. The nature of the motor loss following a lesion of the supranuclear fibers is described on page 1005. In the paralysis resulting from injury to the trunk of the facial nerve all the muscles of the affected side of the face are completely para-

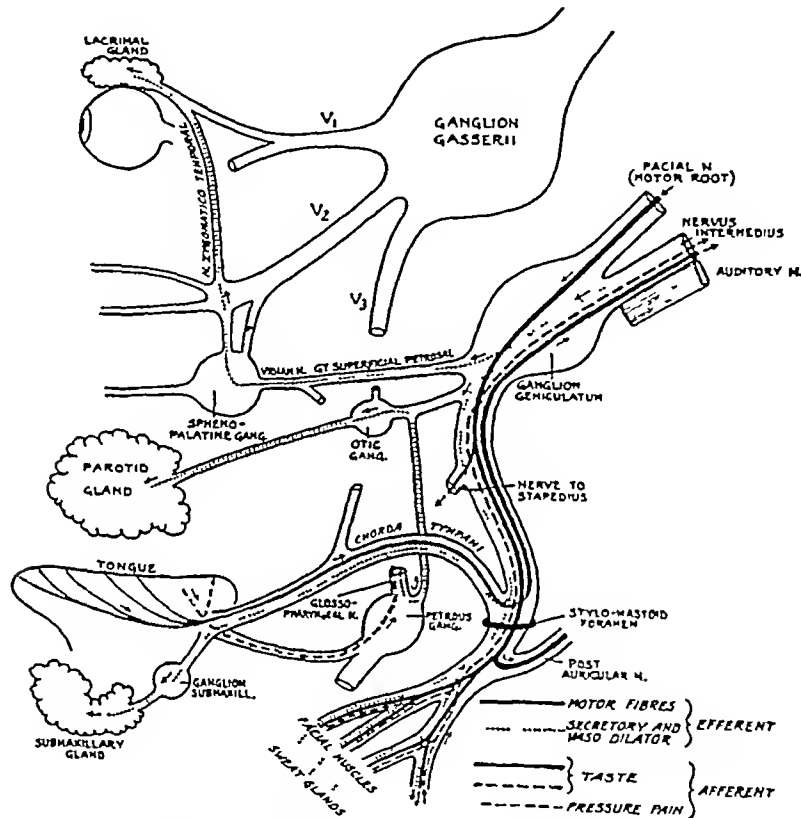


FIG 66 10 Diagram of the facial nerve, afferent and efferent paths, including taste (red) and secretory fibers (dotted lines) (After Purves-Stewart)

prominence in the floor of the 4th ventricle known as the *facial colliculus*

Connections the motor nucleus of the facial nerve receives

(a) Fibers from the corticobulbar tract of the same and of the opposite side.

(b) Fibers from the lateral, trigeminal and medial lemnisci, and from the spinothalamic tracts. Through these connections reflex facial movements in animals may be initiated from various receptive areas of the body.

lyzed. The subject is unable to close the eye owing to paralysis of the orbicularis oculi, or to frown, the eyebrow droops. The mouth is drawn over to the sound side by the unparalyzed muscles and, unlike the paralysis in hemiplegia, the muscles of the affected side do not take part in the facial expression of emotional states, e.g., laughing or crying.

Facial paralysis may result from a lesion involving

(a) The motor nucleus or the intra pontine course of the motor fibers (tumors, hemorrhage, etc.)

(b) The nerve as it crosses the posterior fossa of the skull to reach the internal auditory meatus, as in fractures of the skull or tumors in this situation, the sensory portion and the auditory nerve are commonly involved as well, when loss of the sensation of taste over the anterior two-thirds of the tongue, and deafness on the affected side will result

(c) The nerve in its course through the temporal bone as in fracture of the skull or in otitis media, inflammation of the nerve within the facial canal (aqueduct of Fallopius)—*Bell's palsy*—may occur

(d) The nerve after its emergence from the stylo-mastoid foramen, as it lies behind the angle of the jaw, inflammation, parotid tumors, or accidental injuries may implicate the nerve in this situation

THE GLOSSOPHARYNGEAL NERVE

The glossopharyngeal nerve contains motor, secretory, vasodilator and sensory fibers. The nerve emerges in three or four filaments from the side of the upper part of the medulla in the groove between the olive and the restiform body. The motor fibers are distributed almost entirely to the stylopharyngeus muscle, a few terminate in the circular and longitudinal muscle of the upper part of the pharynx. The secretory and vasodilator fibers (via its tympanic branch, the tympanic plexus, the small superficial petrosal nerve and the otic ganglion) (p 489) supply the parotid gland. The sensory fibers arise from cells in the superior (jugular) and inferior (petrous) ganglia of the glossopharyngeal nerve. The peripheral processes of the cells of the latter ganglion supply the taste buds of the posterior third of the tongue. Those from cells of the superior ganglion mediate ordinary sensations (touch, thermal etc), from this part of the tongue and mucosa of the pharynx and posterior part of the mouth, the central processes (superior ganglion) terminate in the dorsal nucleus of the vagus and (inferior ganglion) in the lower part of the *tractus solitarius*. Fibers (secondary neurons) pass from the latter nucleus and, crossing to the medial fillet of the opposite side, ascend to the thalamus, from where axons of tertiary neurons pass to the cortical area for taste. The motor fibers arise from the upper part of the *nucleus ambiguus* situated in the reticular formation of the medulla. Secretory fibers are the axons of cells lying in the *inferior salivatory* nucleus which lies below the superior nucleus of the same name.

THE VAGUS NERVE

The vagus nerve contains motor, secretory, vasodilator and sensory fibers. The secretory and

vasodilator fibers and the fibers to the involuntary muscle of the bronchi, heart, esophagus, stomach, small intestine, gall-bladder, etc (parasympathetic fibers, p 1095) arise from cells in the *dorsal nucleus of the vagus* (principal autonomic nucleus). This gray mass extends upwards from the lower, closed part of the medulla to beneath the floor of the 4th ventricle at the level of the striae medullares. The voluntary motor fibers arise in close relationship with the motor fibers of the glossopharyngeal, namely, from the cells of the *nucleus ambiguus* lying below the glossopharyngeal neurons. They supply (through the superior laryngeal branch) the cricothyroid and arytenoid muscles of the larynx, and the inferior constrictor of the pharynx. The pharyngeal and recurrent laryngeal branches of the vagus also convey voluntary motor fibers many of which are derived from the bulbar nucleus of the accessory nerve (see below) to the pharyngeal muscles (with the exception of the stylopharyngeus), and to the muscles of the soft palate (except the tensor palati) and larynx (except the cricothyroid). The cell bodies of the sensory fibers lie in the inferior ganglion of the vagus (ganglion nodosum). The peripheral processes of these cells convey impulses from the lungs, heart, larynx, pharynx, esophagus, stomach, small intestine and gall-bladder. They also, through the anterior laryngeal branch, innervate the taste buds of the epiglottis and valleculae (the depressions lying at the sides of the fold running from the epiglottis to the base of the tongue). The taste fibers end centrally by synapsing with cells in the *gustatory nucleus* lying in the upper and medial part of the *tractus solitarius*. These impulses are relayed upwards along the same paths as those conveying other taste impulses. Afferent vagal fibers from visceral structures terminate in the dorsal nucleus. This latter is, therefore, both motor and sensory in function and constitutes an important visceral reflex center. It contains the cardio-inhibitory and vomiting centers.

Vagal afferent filaments travelling in Arnold's nerve mediate the general sensations of the skin lining the external auditory meatus and a small area behind the auricle. Irritation of these fibers may cause reflex cough (p 354). These sensory fibers have their cell stations in the jugular ganglion.

THE ACCESSORY NERVE

The accessory (spinal accessory) nerve is entirely motor and is made up of a bulbar and a

spinal root. The *bulbar* root arises from the lower (caudal) end of the nucleus ambiguus from cells situated below those which give origin to the motor fibers of the vagus. The bulbar fibers join the vagus within and below the jugular foramen, and are distributed, as already mentioned, in the pharyngeal and recurrent branches of the latter nerve; these fibers of the accessory nerve innervate the muscles of the larynx, with the exception of the cricothyroid, the muscles of the pharynx and those of the soft palate, with the exception of the tensor palati (which is supplied by the 5th nerve). The *spinal* root is composed of the axons of a group of cells in the anterior gray column of the cord extending from the 1st to the 4th or 5th cervical segment inclusive. These fibers supply the sternomastoid and trapezius muscles. The spinal part of the nerve exchanges fibers with the bulbar part in the jugular foramen.

THE HYPOGLOSSAL NERVE

The hypoglossal nerve is also purely motor. Its fibers are derived from a nucleus situated near the mid line in the floor of the fourth ventricle and medial to the nucleus ambiguus. It supplies the thyrohyoid, styloglossus, hyoglossus and genioglossus muscles, and the intrinsic muscles of the tongue.

THE PATHWAYS FOR VOLUNTARY MOTOR IMPULSES

The corticospinal fibers convey impulses which bring about voluntary movement from the cerebral cortex (p. 1032) through the white matter of the hemisphere. Converging as they descend they form, together with fibers ascending to the cortex, a fan-like structure known as the *corona radiata*. They enter and stream through the corpus striatum forming here a compact bundle, the *internal capsule*. In the mid-brain they occupy, on each side, the middle two-fifths or so of the base of the cerebral peduncle. In their descent through the pons they become broken up into several smaller bundles by the fibers of the pontine nuclei, but are collected together again at the lower level of the pons and continued downwards as a compact bundle or column, called the *pyramid*, on the ventral aspect of the medulla, and in contact at the midline with its fellow of the opposite side. It is from these structures that the name "pyramidal", alternately applied to the corticospinal tracts, has been derived. At the lower limit of the medulla each pyramidal tract divides into two bundles of unequal size. The larger of these (crossed pyramidal tract) contains four-fifths or more of the total

number of fibers and decussates with the corresponding bundle of the opposite side, the smaller bundle descends uncrossed (direct pyramidal tract, see p. 994).

The corticospinal or pyramidal fibers are not derived entirely or even predominantly from the cells of Betz, of the motor area as was once believed, for the numbers of fibers in the pyramids exceeds by some thirty times the number of Betz cells, *only between 2 and 3 per cent of the corticospinal fibers are the axons of the cells of Betz*.²

The fibers derived from the Betz cells are of much larger diameter (10-20 micra), more heavily

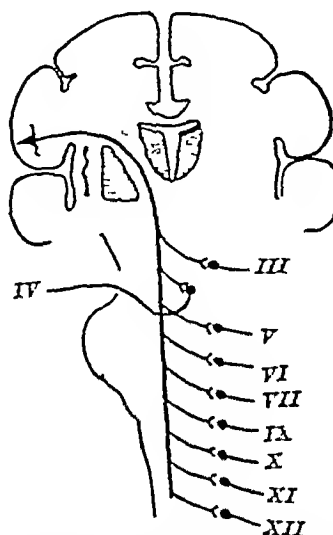


FIG. 66-11. Diagram of the motor nuclei of the cranial nerves and their cortical connection (in part after Ranson).

myelinated and of more rapid conduction than the more numerous fibers, which suggests that they are concerned with the finer muscular movements, and the smaller fibers with the grosser more general movements, or with muscle tone. Lassek found that over 90 per cent of these small fibers had a diameter of only from 1 to 4 microns, and less than 2 per cent more than 10 microns. Some of these fibers are myelinated, some unmyelinated. Their sites of origin are not altogether known, but some are derived from the motor area, others from the premotor cortex, and yet others from the postcentral parietal, and probably other regions of the cortex.

² According to Lassek, the number of Betz cells in the human motor cortex of one hemisphere is around 34,000 whereas the number of fibers in the corresponding pyramid is over 1,000,000.

The manner in which the corticospinal fibers are connected with the cells of the spinal cord, and the convergence upon the latter of impulses from higher levels, e g, red nucleus, superior colliculus, labyrinth, cerebellum, etc, have been touched upon in the account of the fiber tracts in the cord. The motor neuron as a part of the reflex arc has been dealt with in chapter 64.

The fibers passing from the lower part of the motor area of the cortex to the cranial nuclei constitute the *corticobulbar tract*. This bears pre-

responding areas of the cerebral cortex to the nuclei of the pons.

- (b) *Corticobulbar tract*, from the frontal lobe to the red nucleus.
- (c) *Frontothalamic tract* from the frontal lobe, mainly from area 9 of the prefrontal area to the thalamus.
- (d) *Corticostriatal fibers* from the premotor area of the cortex to the caudate nucleus and the globus pallidus.
- (e) *Occipitotectal fibers* from the visual cortex to the superior colliculus.

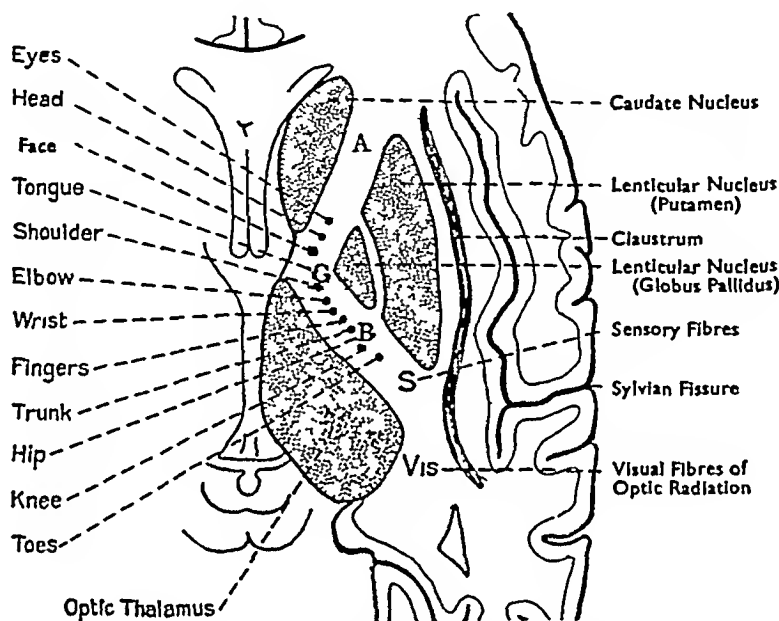


FIG 66 12 Horizontal section of right hemisphere, showing corticobulbar and corticospinal fibers, and the location of the visual and sensory fibers. A, anterior limb of capsule, B, posterior limb, G, genu.

cisely the same relationship to the cells of the cranial motor nuclei as does the corticospinal tract to the anterior horn cells of the cord. The corticobulbar fibers, however, cross at various levels throughout the brain stem. (See fig 66 11.)

THE INTERNAL CAPSULE

In a horizontal section of the cerebrum, the internal capsule is seen as a compact band of white matter lying between the thalamus and caudate nucleus on its inner aspect and the lentiform nucleus on its lateral aspect (cf fig 66 12). In addition to the corticospinal and corticobulbar tracts the following are the principal tracts which compose it.

I Descending

- (a) *Corticopontine tracts* consisting of four bundles of fibers, *frontopontine*, *temporopontine*, *parietopontine* and *occipitopontine* from the cor-

II Ascending

- (a) *Thalamic radiations* consist of four bundles ascending from the thalamus, *anterior*, *superior*, *posterior* and *inferior*, which connect the thalamus with various parts of the cerebral cortex. The anterior group terminate in prefrontal areas, and the superior group in the somesthetic area (postcentral gyrus).
- (b) *Auditory radiations*—fibers ascending from the medial geniculate body to the temporal cortex (geniculo-temporal fibers).
- (c) *Optic radiation*—fibers ascending from the lateral geniculate body to the visual area in the occipital cortex (geniculocalcarine) fibers.
- (d) *Rubro cortical fibers* from the red nucleus to the cortex.

The internal capsule is bent in the horizontal plane to form a convexity directly medially. The region of the angle so produced is called the *knee* (*genu*), the portions in front and behind this are

known, respectively, as the *anterior* and *posterior limbs*. The shorter anterior limb lies between the caudate nucleus and the thalamus, the posterior limb between the lentiform nucleus and the thalamus. The extension backwards of the posterior limb is known as the *retrolenticular part*.

The anterior limb is occupied by the *fronto pontine* fibers and the *anterior thalamic radiation* (thalamofrontal fibers), the *frontothalamic tract*, *corticostriatal* and *corticorubral tracts*. In the anterior three fifths of the posterior limb are transmitted the *corticobulbar* and *corticospinal* (pyramidal) fibers. The fibers of these two tracts are segregated into groups in accordance with the muscles which they govern, those carrying impulses for the eye muscles, the muscles of the tongue, and *face* and *upper* and *lower limbs arranged in this order from before backwards*, and those transmitting impulses destined for the proximal part of a limb are placed in front of those for the more distal muscles. The remaining part of the posterior limb, and the retrolenticular part carry in antero posterior order the *superior* and *posterior thalamic radiations*, the *auditory* (*geniculotemporal*), and the *inferior thalamic radiations*, *corticothalamic*, *temporopontine* and *parietopontine* fibers, and the *optic* (*geniculo occipital*) radiations.

The internal capsule is supplied with blood through the lateral striate branches of the middle cerebral artery. One of these (lenticulo striate), somewhat larger than the others, was called by Charcot the "artery of cerebral hemorrhage" (see below).

A COMPARISON OF THE EFFECTS PRODUCED BY INJURY OF SUPRASPINAL PATHS WITH THOSE OF THE SPINAL CENTERS OR PERIPHERAL NERVES (MOTONEURONS)

It has been the teaching for many years that the voluntary motor pathway comprised two links, the upper of which was a corticospinal (or corticobulbar) fiber, and the lower a ventral horn cell of the spinal end, i.e., the spinal motoneuron. The former was called the *upper motor neuron*, the latter the *lower motor neuron*, which is also the final common path to the skeletal muscles (ch. 64). The clinical manifestations attributed to the "upper neuron" were characteristic of it, and distinct from those of the other. The so called "upper motor neuron" is now recognized as consisting of two links in parallel, the corticospinal (pyramidal) tract and an extrapyramidal pathway. The clinical manifestations will depend upon

whether the one or other component or both together are interrupted.

The clinical signs which were supposed to be due to interruption of the corticospinal fibers alone, and, therefore, characteristic of it, are listed below.

- (1) Paralysis of muscle groups which ordinarily carry out a certain *coordinated movement* and not of individual muscles.
- (2) Hypertonia (spasticity) of the paralyzed muscles.
- (3) Exaggeration of tendon reflexes, knee jerk increased, ankle clonus present.
- (4) Positive Babinski response (p. 1013).
- (5) No permanent wasting, except as a result of prolonged disuse.

(6) Absence of the reaction of *degeneration*.

It has been found, however, that some of these effects, e.g., hypertonus, and increased tendon reflexes, are not seen in an injury confined to the pyramidal fibers and that the complete picture appears only when the extrapyramidal fibers connecting the premotor cortex (area 6) with the spinal cord are involved as well. A lesion of the extrapyramidal system alone, e.g., ablation of the premotor area (chapter 68) causes hypertonus, moderate increase in the tendon reflexes, a positive Rossolimo reflex and lateral deviation or "fanning" of the toes, but the plantar response is of the normal flexor type, i.e., the Babinski ("extensor" response) is absent. Injury to both systems, as in cerebral hemorrhage, causes spasticity, i.e., a type of hypertonus resembling that of the decerebrate preparation, exaggeration of the tendon reflexes and positive Babinski and Rossolimo reflexes. There is also "fanning" of the toes (see also p. 1014).

CEREBRAL HEMORRHAGE

The corticospinal pathway is most commonly injured as the result of a vascular lesion, hemorrhage or thrombosis (usually of a lenticulo-striate branch of the middle cerebral artery) within the internal capsule. Cerebral hemorrhage of this nature is commonly called *apoplexy* or a "stroke". The lesion is usually unilateral and its effects involve the muscles on the opposite side of the body. The unilateral paralysis of the limbs is called *hemiplegia*. After the state of shock, which lasts for a variable number of days following the injury, has passed off, the following features are observed.

- (1) *Loss or impairment of voluntary power*, involving the muscles of limbs, trunk and face of

the opposite side is evident. There is often little or no actual paralysis, merely a muscular weakness (paresis). The finer movements, such as those of the hands and fingers are especially affected. The volitional movements of the lower part of the face are involved to a greater degree than those of the upper, e.g., raising the eyebrows and closure of the eyelids, the reason being, very probably, that the part of the facial nucleus governing the latter movements receives fibers from both hemispheres. When voluntary efforts are made to move the lower part of the face, as in showing the teeth, or pursing the lips, marked impairment of muscular power may be evident, yet emotional expressions, e.g., laughing, smiling or crying, though involving the same muscles may show little departure from the normal. For example, the patient, though unable to raise the corner of his mouth when asked, may smile naturally a moment later. The impulses which bring about movements of the facial muscles in these more automatic movements of facial expression apparently travel by pathways other than pyramidal, and have escaped interruption.

(2) *Movements of the jaw, tongue and soft palate* on the opposite side to the pyramidal lesion show some weakness but since they are bilaterally innervated this, as a rule, is not pronounced. There may be some deviation of the jaw upon opening the mouth, and of the tongue when protruded, to the hemiplegic side. The soft palate is lower on the hemiplegic side and the uvula tends to be drawn away from this side (i.e., toward the side of the cerebral lesion).

(3) *Hypertonia of the muscles—spasticity*. This involves on the paralyzed side the flexors of the elbow, wrist and fingers, adductors of the shoulder and pronators of the forearm. In the lower limb the extensors of the hip and knee, the adductors of the hip and the plantar flexors of the ankle are hypertonic. The attitude resulting from this distribution of the hypertonia is similar in character to the decerebrate rigidity of animals (p. 964). The muscles show the lengthening and shortening reactions of the decerebrate preparation. The hypertonus has an elastic quality, this, as well as its unequal distribution, distinguishes it from the "lead pipe" type of rigidity seen when the lesion is confined to the extrapyramidal neurons.

In cerebral hemorrhage the injury is not confined to the pyramidal fibers, but involves as well extrapyramidal fibers which descend with the latter in

the internal capsule. All the signs listed on page 1004 are, therefore, present. The muscles do not waste and the electrical reactions (p. 914) are normal. (See also the premotor area of the cortex, ch. 68.)

(4) *Reflexes*. The tendon jerks at knee, ankle and elbow are exaggerated on the hemiplegic side and there may be ankle clonus. The plantar response is "extensor" in type⁴; the abdominal and cremasteric reflexes (p. 1014) are reduced or abolished.

The corticospinal neuron may be involved in injury or disease (hemorrhage or thrombosis, tumor, etc.) in any part of its course from the cerebral cortex to the spinal centers. The effects, though showing the general features just described, e.g., loss or impairment of muscular power, hypertonia and exaggerated reflexes, present certain features peculiar to the level of the lesion. When the latter implicates the tract where, as in the internal capsule, the fibers form a compact bundle the peripheral effects are likely to be more widespread, i.e., to be *hemiplegic*, than when the motor area of the cortex or the corona radiata is involved. Then the paralysis tends to be more localized and may affect only one limb—*monoplegia*.⁵ Again, when the lesion interrupts the fibers anywhere above the nuclei of the motor cranial nerves (supranuclear lesion) the muscles innervated by the latter will, of course, be involved as described above with the muscles of the limbs and trunk. The fibers from the cortex of the motor nuclei cross at different levels but all above the decussation of the corticospinal fibers. If, therefore, a lesion involves the pyramidal pathway below, say, the nucleus for the seventh cranial nerve but above the decussation of the corticospinal fibers the paralysis will involve the limbs and trunk of the opposite side and perhaps those of the tongue, soft palate and throat, but the facial muscles will escape. When the lesion involves the facial nucleus itself together with the pyramidal tract, *crossed* or *alternate hemiplegia* results. This is known as the Millard-Gubler syndrome and consists of paralysis of the limbs and trunk of the spastic type on the side opposite to the lesion, together with the facial paralysis of the lower neuron type on the side of the lesion. Similarly, the hypoglossal nucleus may be involved together

⁴ This is in reality a flexor reflex (see p. 1013).

⁵ Paralysis involving all four limbs is called *quadriplegia*, if only the upper limbs are paralyzed the term *diplegia* is used, or *paraplegia* when only the lower limbs are paralyzed.

with the pyramidal tract above its decussation. Wasting and paralysis of the tongue muscles on the same side as the lesion, and spastic paralysis of the limbs on the opposite side will result. Again, the glossopharyngeal nucleus may be implicated with consequent paralysis of the stylopharyngeus

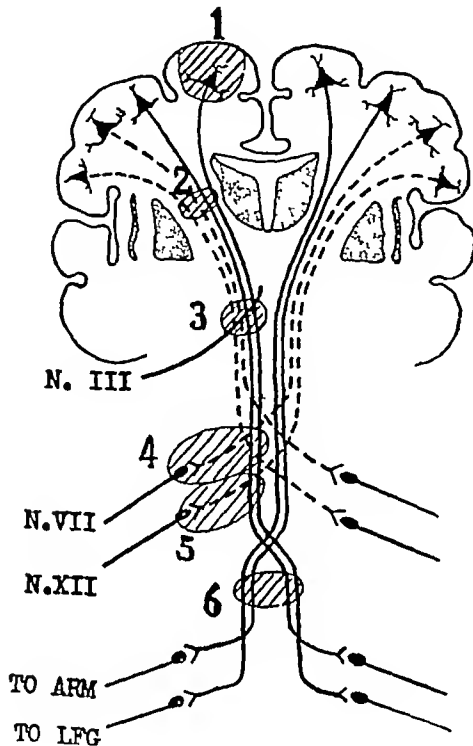


FIG 66 13 Diagram illustrating the effects of lesions at various levels. 1, lesion of cortex causing monoplegia or hemiplegia, depending upon its extent, 2, lesion of internal capsule, hemiplegia, 3, lesion of mid brain involving pyramidal fibers and nucleus or tract of the third nerve (Weber's paralysis), 4, lesion in pons destroying pyramidal fibers and nucleus of facial nerve, crossed hemiplegia with homolateral facial paralysis, 5, lesion in medulla involving pyramidal fibers and the hypoglossal nucleus, crossed hemiplegia and homolateral paralysis of lingual muscles, 6, transection of cord, paraplegia (in part from Villiger and from Ran son)

and loss of sensation over the posterior third of the tongue. The oculomotor nucleus, since it lies in close proximity to the pyramidal tract, is also sometimes involved with the latter. The level of the lesion would then be indicated by paralysis of the ocular muscles innervated by the third nerve on the side of the lesion and hemiplegia on the opposite side. This is known as *Weber's paralysis* (fig 66 13)

It is evident that the paralysis is more likely to be bilateral if the lesion is situated in a region where, as in the brain stem or spinal cord, the corticospinal tracts of the two sides are approximated, than if it is located at a level where these tracts are widely separated, as in the corona radiata or internal capsule, then unilateral effects are usual. Injury to the crossed pyramidal tract on one side of the cord, should it occur, will result in paralysis of the muscles on the same side. Complete interruption of both pathways in the cord will of course result in paralysis on both sides of the body below the level of the lesion—*paraplegia* (p 1011)

LESIONS OF THE SPINAL MOTONEURON (LOWER MOTOR NEURON)

Destruction of the motor cells situated in the anterior horns of the cord or of the motor nuclei of the cranial nerves, of the anterior spinal nerve roots or of the peripheral nerve trunks, results in complete paralysis of the muscle fibers supplied by the injured motoneurons. The paralysis is of the *flaccid type*, of which the following are the main clinical features

- (1) Paralysis not of movements essentially, but of individual muscles or even parts of muscles (see motor units, ch 66)
- (2) Hypotonus, flaccidity
- (3) Loss of tendon reflexes
- (4) Atrophy (wasting) and degeneration of muscles
- (5) Electrical reactions of degeneration (p 914)
- (6) Plantar response if present is of the normal flexor type

The paralysis does not, as in lesions of the upper neuron, show a hemiplegic distribution. When the lesion involves the anterior horn cells as in anterior poliomyelitis, or the anterior roots of the spinal nerves, the paralysis assumes a segmental pattern. Individual muscles or scattered groups of muscles are commonly involved. Injury to the anterior horns usually results in a bilateral paralysis, though different muscles on the two sides of the body may be affected. Furthermore, since a given muscle receives its innervation from more than one spinal segment, and conversely a single spinal segment supplies parts of several muscles (p 954), a localized lesion of the cord may result in the paralysis of fibers scattered throughout several muscles, leaving other fibers in the same muscles intact. Weakness rather than complete paralysis of certain muscles will result. Even when

the lesion involves several spinal segments, the muscles innervated by those at the upper and lower limits of the spinal region involved, will show partial paralysis, since they will still be supplied by undamaged nerve fibers from adjacent healthy segments

The entrance of the ventral roots into the formation of the brachial and lumbar plexuses is accompanied by a regrouping of their constituent fibers which are continued into the various peripheral nerve trunks. A *lesion of a peripheral nerve*, therefore, produces a paralysis which is non-segmental in character. It corresponds to the distribution of the muscular branches of the nerve.

Cutaneous sensory loss will show a segmental pattern when the posterior roots of the spinal nerves are involved, but when a peripheral nerve trunk is implicated the distribution of the sensory defect will conform to that of the cutaneous branches of the nerve (see also figure 66.3). Injury to the ascending tracts in the cord will, of course, cause sensory loss below the level of the lesion (p. 993) and will vary in kind according to the particular fibers which have been interrupted.

FUNCTIONAL, MORPHOLOGICAL AND CHEMICAL CHANGES IN DENERVATED MUSCLE

A lesion which involves any part of the lower motor neuron—anterior horn cell—such as may result from anterior poliomyelitis or peripheral nerve injury, is followed by degenerative changes which extend to and include the nerve terminals within the muscle. A muscle thus completely denervated exhibits constant, fine, rapid, rhythmical contractions. This so-called fibrillation of the denervated muscle does not appear until several days after the abnormal electrical reactions (reaction of degeneration, p. 914) have already developed and complete degeneration of the nerve has occurred. The muscle fibers contract asynchronously, the contractions involve only a part of the length (0.5 to 1 mm) of the fiber and give rise to small irregular action potentials. The fibrillation can sometimes, though not commonly, be seen through the skin. Having once appeared, fibrillation persists for a year or more and until the contractile elements have undergone complete atrophy. According to Denny-Brown, a reduction in tension developed by the muscle when stimulated maximally can be detected within 2 minutes after section of the nerve, but other investigators have been unable to observe any loss of contractile force until after a much longer interval (30–40 hours). An outstanding functional effect of denervation is the great increase in the sensitivity of the muscle to the intravenous or intra-arterial injection of acetylcholine.

Normal muscle is excited by the intra-arterial injection of from 0.2 to 2.0 mg of acetylcholine, whereas denervated muscle responds to a dose of from 0.002 to 0.02 γ . It is generally believed that this hypersensitivity to acetylcholine is responsible for the Vulpian and Sherrington effects, and most probably also for the "fright reaction" described by Bender (p. 1101). The denervated muscle is also hyperexcitable, though to a less pronounced degree, to potassium chloride.

The denervated muscle soon commences to atrophy, reduction in its bulk becoming apparent within a few days after section of the nerve. Atrophy progresses rapidly and is followed by degeneration of the contractile elements. Microscopic changes in the muscle fibers consist of swelling and vesiculation of the nuclei, a reduction in sarcoplasm, and fading, followed by disappearance of the striations of the myofibrillae. Ultimately all contractile tissue disappears and is replaced by fibrous tissue and fat. As these morphological changes are occurring, the muscle gradually loses its plastic or ductile quality, flexors and adductors shorten and become more or less "set" in their new positions. The antagonistic muscles lengthen adaptively. This state of the muscles is called *contracture*.

The chemical changes in denervated muscle comprise a reduction in glycogen, phosphocreatine and adenosinetriphosphate, but they do not become pronounced until the onset of fibrillation. The breakdown and subsequent resynthesis of glycogen by the denervated muscle during a work period is normal for a time following nerve section but when fibrillation supervenes, the ability of the muscle to restore its glycogen after contraction is greatly impaired. The atrophying muscle shows a large increase in calcium, a smaller increase in chloride, and a decrease in potassium. The changes in chloride and potassium can be accounted for by the reduction in muscle mass and its replacement by interstitial tissue and fluid. The increase in calcium is too great, however, to be explained entirely in this way.

The cause of fibrillation in denervated muscle is unknown, but it is not improbable that it is a manifestation of its hypersensitivity to acetylcholine. The observation of Magladery and Solandt that quinidine, which abolishes acetylcholine hypersensitivity, also suppresses fibrillation is highly suggestive, yet as Dale and his associates have shown, direct stimulation of denervated muscle does not cause the release of acetylcholine. It follows, therefore, that if the stimulating effect of this agent is responsible for fibrillation it must be carried to the muscle in the blood stream. The increased sensitivity to acetylcholine may be due to a lack of cholinesterase for, according to Marnay and Nachmansohn, this enzyme, which normally is concentrated in the neighborhood of the nerve endings in skeletal muscle, disappears after denervation. The lack of cholinesterase might also explain the failure of eserine to cause anything more than a transient augmentation of fibrillation.

The atrophy following nerve degeneration has not been explained satisfactorily. Langley and Kato believed that the ceaseless activity of the fibrillating muscle was the primary cause, thus attributing it to overwork and exhaustion. Though other investigators still incline to this view it has not received general acceptance. One reason for questioning this explanation is that the tension developed by the fibrillating muscle is less than that of the ordinary tonic contraction of normal muscle, yet the oxygen consumption of the denervated muscle is greater than that of normal controls. One would not expect atrophy to result under such conditions. There is little evidence for the earlier belief in trophic nerves in the original meaning

extent unless the lesion involves the cervical or lumbar enlargements (fig 66 14)

II Vasomotor paralysis

As a result of the interruption of pathways in the lateral funiculus connecting the vasomotor center with the spinal sympathetic centers (intermediolateral cell column) dilatation of the cutaneous vessels below and on the side of the lesion occurs. The skin of the paralyzed side is at first redder and warmer than normally, later it becomes cyanosed and cold.

III Sensory changes

(a) *Sense of position and movement and of tactile discrimination* (compass test), *stereognosis* and the *perception of vibration*, being mediated by fibers of the posterior columns of the cord which do not cross to the opposite side, are lost below the level of and on the same side as the lesion. The loss of the sensations from the muscles and joints causes ataxia on the side of the lesion. The non sensory impulses from the muscles since they pass to the cerebellum by paths (direct and indirect spino cerebellar) containing both crossed and uncrossed fibers are not completely interrupted on either side. The partial blocking of these impulses, however, contributes to the ataxia.

(b) *Painful and thermal sensations* are conveyed by fibers which cross shortly after their entrance into the cord (lateral spinothalamic tract). There is therefore complete analgesia and thermoanesthesia on the side opposite to the lesion, except for a narrow band immediately below the level of the injury in which these sensations are retained, for the fibers over which the impulses travel ascend for a segment or so before crossing to the injured side and, thus escaping interruption, proceed upwards above the lesion. Destruction of fibers of the posterior roots as they enter the cord causes a narrow band of complete anesthesia and analgesia at the level of the lesion.

(c) *Touch and tactile localization*. It has been pointed out (p 993) that fibers mediating these sensations cross at different levels in the cord. Therefore, a unilateral lesion of the cord interrupts fibers which have crossed from the opposite uninjured side of the body as well as those ascending from the same side, i.e., uncrossed fibers of the posterior columns. These sensations on the side of the lesion are still mediated, however, by fibers which have escaped interruption through having crossed below the lesion ascend in the ventral

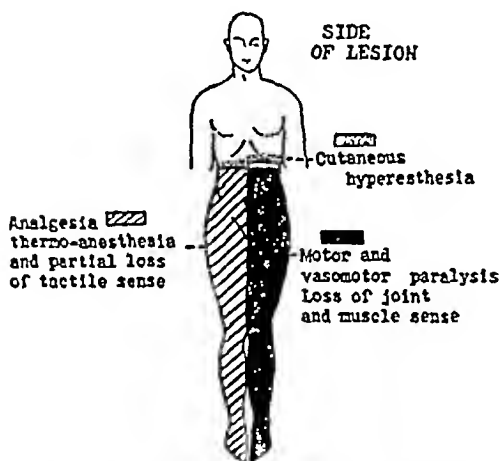


FIG 66 14 Diagram showing effects following hemisection of the cord in the lower thoracic region. Narrow white band at level of lesion indicates total anesthesia and analgesia.

of the term, namely, that certain nerve fibers exerted a specific nutritional or trophic control over the muscle.

HEMISECTION OF THE CORD BROWN-SEQUARD'S PARALYSIS

A lesion which completely interrupts the continuity of one-half of the cord causes the following disturbances

I Motor paralysis

Upper neuron type of paralysis of the muscles below the level of the lesion and on the same side occurs as a result of the severance of the cortico-spinal tracts. The muscles supplied by the segments actually involved in the lesion will, as a result of destruction of the anterior horn cells, exhibit a flaccid paralysis (lower motor neuron type). The flaccid paralysis is, however, small in

(anterior) spinothalamic tract of the uninjured side, while those from the side opposite to the lesion are carried by the fibers of the posterior columns of that side. The sensations of touch and tactile localization, therefore, are not lost on either side of the body as a result of a unilateral lesion of the cord, though some impairment on the same side as the lesion may be detected.

Finally, there may be a narrow zone of hyperalgesia and hyperesthesia on both sides due to the irritation of the sensory fibers within the cord.

SYRINGOMYELIA

This commences as a proliferation of neuroglia (gliosis) in the spinal gray matter. The newly formed tissue breaks down with the formation of cavities filled with a gelatinous material. The process may extend for considerable distances up and down the cord. The starting point of the disease is most frequently the base of the posterior gray column. The lower cervical and upper thoracic regions are most commonly affected. The disease may extend into the bulb (syringobulbia) when signs of implication of the cranial nuclei appear.

Dissociated sensory loss and muscular weakness with a segmental distribution are among the characteristic features of the disease. Owing to the site of the changes, the fibers mediating pain, heat and cold, where they make connections with the cells of the posterior horns, are first destroyed. If the lesion is unilateral, thermoanesthesia and analgesia of the skin supplied by, and on the same side as the diseased segments result. Other sensations, e.g., touch, muscle sense, are unaffected until later in the disease. Involvement of the spinal nucleus of the trigeminal results in the characteristic sensory loss over the face, the areas supplied by the ophthalmic, maxillary and mandibular divisions being affected successively in this order (p. 998). When both dorsal horns are involved, or as a result of expansion of the cavity the anterior gray commissure is pressed upon and the fibers injured at their point of crossing, the thermoanesthesia and analgesia are bilateral. The muscles innervated by the anterior horn cells of the diseased segments become weak and wasted, these effects often appear first in the small muscles of the hand. So long as the disease does not cause injury to the white matter, the sensory and motor effects are limited and have a segmental distribution. Involvement of the corticospinal tracts will result in spasticity below the level of the lesion, while pressure upon the spinothalamic tracts and dorsal columns will be followed by sensory loss below the level of the lesion. The degree of sensory loss, its type and distribution and whether homolateral or contralateral vary according to the extent to which the individual tracts are affected by the disease. As a result of the loss of the protective

sensations (pain and temperature), subjects of syringomyelia are prone to suffer injuries which, not being perceived at the time, are neglected and ultimately lead to serious lesions, e.g., painless disorganization of a joint (arthropathy, Charcot's joint). Involvement of the sympathetic centers in the lateral columns leads to vasomotor disturbances, excessive sweating or absence of sweating, cyanosis, etc. So-called trophic disturbances, e.g., ulcers, whitlows, gangrene, etc., are largely the result of vasomotor abnormalities and the loss of sensation which, as just mentioned, permits an injury, trivial perhaps at first, to be disregarded.

SUBACUTE COMBINED DEGENERATION OF THE CORD

This condition involves the white matter of the cord and is almost always associated with pernicious anemia. The degenerative process consists of a breakdown of the myelin sheaths, subsequent destruction of the axons and their replacement by newly formed glial tissue. The changes are most pronounced in the corticospinal and cerebellar tracts and in the posterior columns. The spinothalamic tracts, as a rule, are involved to a less degree. The chief clinical features are, therefore: (a) muscular weakness and spasticity, (b) impairment of the sense of position and passive movement of the limbs with consequent ataxia and a positive Romberg sign, (c) loss or impairment of sensitivity to touch, pressure, localization, spatial discrimination, vibration, pain and temperature. The relative extent to which these sensations are lost is variable. The cutaneous sensory loss is at first over the distal regions of the extremities—hands and wrists, feet and ankles. Paresthesias (tingling, pricking and burning sensations) frequently precede the sensory loss. (d) The reflexes vary. The tendon jerks are frequently exaggerated as a result of involvement of the corticospinal tracts, but they may disappear as the disease progresses. The plantar response is usually of the "extensor" type, especially in the later stages.

The severity of the nervous manifestations do not always run parallel with that of the blood picture, the neurological features may be pronounced though the anemia is of mild degree, or vice versa. Subacute combined degeneration of the cord, though associated with pernicious anemia, is not due to the anemia itself. This is shown by the well recognized fact that the blood picture may be restored to normal by the administration of folic acid, but the neurological condition progresses unchecked, whereas the latter is arrested by the hematonic principle (vitamin B₁₂). See also ch. 9.

TABES DORSALIS, LOCOMOTOR ATAXIA

In this condition, which is the result of syphilis, the fibers of the dorsal roots after their entry into the cord (i.e., the central processes of the primary sensory neurons) are attacked. The ganglion cell bodies of

the dorsal roots, as a rule, are not affected. The essential lesion within the cord, therefore, involves the entrance zone of the lateral division of the dorsal roots (dorsolateral fasciculus of Lissauer), and the dorsal fasciculi (gracilis and cuneatus). The endogenous fibers of the cord escape, but although it is those tracts composed of exogenous fibers which are specifically attacked by the disease, the functions of the cerebellar and spinothalamic tracts are also seriously disturbed as a result of degeneration of the primary neurons leading to them. The descending tracts remain as a rule practically unaffected. The reason for the selective destruction of the exogenous fibers of the cord is unknown, possibly it is compression of the fibers by proliferation or inflammatory swelling of the meninges at the point of entrance of the posterior root. The central fibers of the trigeminal nerve which are homologous with the fibers of the posterior spinal nerve roots may also be implicated.

The sensory changes are those which might be expected to result from a gradual degeneration of dorsal root fibers. Somewhat similar effects are produced in monkeys by section of the posterior nerve roots.

The chief manifestations are as follows: (a) During the degenerative process paresthesias of various types, hyperesthesia and stabbing pains are common. (b) Impairment, or loss to a variable degree, of all forms of sensation follows. Loss of the rapid conduction component of deep pain occurs early. Loss of the sense of position and of passive movement, and blockage of afferent cerebellar impulses result in marked incoördination of the muscles—ataxia. Movements are jerky, exaggerated and imperfectly controlled, the subject being unable to move his hands and feet in the desired direction or to assume a given position at will, e.g., bringing his finger to the nose or placing the heel of one foot upon the toes of the other. The gait is ataxic, the feet are kept wide apart, raised unnecessarily high and brought down in a stamping fashion, the patient may learn to overcome this tendency by shuffling.

Involvement of sensory cranial fibers causes corresponding incoördinate actions of the facial, ocular and lingual muscles. When standing with the eyes closed the patient, being thus deprived of an important aid in maintaining his equilibrium, tends to sway and may fall (Romberg's sign). (c) The interruption of the pathways for proprioceptive impulses from the skeletal muscles (p. 942) results in extreme *hypotonia*, but there is no true paralysis. (d) The *tendon reflexes* are abolished as a result of the destruction of the afferent limb of the proprioceptive reflex arc. The abdominal reflexes are present. (e) "*Lightning pains*" and *trophic disturbances*. The former are severe stabbing paroxysmal pains usually localized to an area supplied by one or more spinal segments. Vasodilatation, small hemorrhages or herpes zoster may occur in the painful area. These vascular and cutaneous effects have been attributed to antidromic impulses reaching the periphery via sensory

fibers. The skin of the affected area may break down with the formation of so-called trophic ulcers. Painless destruction of joints (Charcot's joint) is not uncommon in tabes. The loss of the sense of pain which causes the patient to suffer injuries of which he is unaware, the extreme hypotonia of the muscles which normally support the joint, and the vascular disturbances resulting from damage to autonomic fibers combine to produce such joint conditions. They are usually classed among the trophic disturbances, the term implying that the interruption of trophic impulses is responsible. However, the existence of true trophic fibers, i.e., specific fibers which preside over the nutrition of the peripheral tissues, is questionable. (f) *Tabetic crises*. These are apparently the result of the involvement of afferent autonomic fibers which enter the cord by the dorsal roots. They consist of paroxysmal attacks of pain and functional disturbances in one or other of the viscera. Gastric crises are the commonest. They consist of severe epigastric pain and vomiting. *Rectal crises* consisting of pain in and increased activity of the rectum, *vesical crises* with bladder pain and difficult urination or *laryngeal crises*, in which spasm of the adductors of the larynx with dyspnea may occur. (g) *Ocular signs*. The pupils are as a rule constricted and often unequal. The Argyll Robertson pupil (p. 1177) in which the reflex to accommodation is retained but the reaction to light is lost, is a characteristic ocular feature of tabes. The pupil also frequently fails to respond by dilatation to stimulation of the skin of the neck (cilio-spinal reflex). The loss of this reflex is usually attributed to degeneration of the central sympathetic pathway through which the dilator pupillae muscle is innervated (p. 1175). Some drooping of the upper lid (ptosis) may also result from the blockage of sympathetic pathways which normally transmit impulses to the smooth muscle in this situation, compensatory contraction of the frontalis muscle with wrinkling of the skin of the forehead results. Damage to the fibers of the 3rd, 4th or 6th nerves results in paralysis of the ocular muscles, the external rectus most commonly. Squint and double vision (diplopia (p. 1184)) are consequences. Primary optic atrophy occurs.

Disseminated (multiple) sclerosis. This is a chronic disease of the nervous system characterized anatomically by the occurrence of small patches of demyelination followed by overgrowth of glial tissue throughout the white substance of the brain (especially in the regions beneath the lateral ventricles), and spinal cord, as well as in various cranial nerves. The lesions in the cord are most numerous in the corticospinal tracts and less pronounced, as a rule, in the posterior columns. Clinically, the disease is marked by an insidious onset, irregular course, with remissions and relapses of unpredictable duration, and a great variability of signs and symptoms between individual cases. Progressive weakness of the muscles of the legs, with the characteristics of an upper neuron lesion, e.g., increased

tendon jerks, "extensor" plantar response (Babinski) and loss of abdominal reflexes result from the involvement of the pyramidal tracts, or paresthesias, loss of vibration and muscle senses from lesions in the posterior columns. The loss of tactile and other sensations is variable. Impairment of vision, or even blindness, may result from involvement of the optic nerve, chiasma or tract. Nystagmus, slurring or staccato speech may occur from the disease affecting other cranial nerves or tracts, and an intention tremor (Charcot's triad) appear as the disease advances. Muscular weakness progresses inexorably to paraplegia in flexion. Often a peculiar feature of the disease is a psychological one—the extremely cheerful attitude of the patient in the face of the complete hopelessness of his condition.

COMPLETE TRANSVERSE DIVISION OF THE CORD

A sudden, or rapidly progressive, complete interruption of the continuity of the cord may result from injury (e.g., gunshot wound, fracture-dislocation of the spine, etc.) or from acute inflammation (e.g., transverse myelitis). Immediate and complete loss of voluntary power below the level of the lesion results. Paralysis of both lower limbs resulting from this or any other nervous lesion is spoken of as *paraplegia*. Complete division of the cord in the lower cervical region will result in paralysis of all four limbs, *quadriplegia*. A lesion of this nature in the upper cervical region is of course rapidly fatal since the diaphragm and other respiratory muscles are isolated from the respiratory center. The subsequent history of the subject of a complete transverse spinal lesion has been divided into three stages by Riddock.

I Stage of spinal shock Immediately following the injury there are complete loss of visceral and somatic sensations, and flaccid paralysis below the level of the lesion. The skeletal muscles are quite toneless. The tendon jerks, plantar response and abdominal reflexes are abolished. The cremasteric and bulbocavernosus reflexes, though absent as a rule, may at times be elicited. The anal reflex is present. A zone of heightened sensitivity (hyperesthesia) immediately above the level of the lesion is present and spontaneous pains in this region, or a feeling of tightness encircling the body may be experienced by the patient. There is retention of urine and feces due to tonic contraction of the sphincters. This stage is analogous to the state of spinal shock in lower animals but is much more severe and prolonged (ch. 67). In man spinal shock lasts for from one to three weeks.

II Stage of reflex activity—paraplegia in flexion

Recovery of the isolated spinal centers commences, the flexor muscles gradually regain a part of their lost tone. The first reflex to appear is the so-called extensor plantar response or sign of Babinski (see p. 1013). At this time the abdominal (rectus muscles) and other superficial reflexes can also usually be evoked with ease. Later the tone of the extensor muscles is restored to some extent and the tendon jerks reappear. The knee jerk returns usually in from 3 to 7 weeks after the injury and may become somewhat greater than normal, but shows certain abnormal features. The extension of the knee is not maintained in the normal way for a brief period, instead, the quadriceps after its contraction relaxes suddenly again and allows the leg to fall like a dead weight. The ankle jerks are weak or may not appear at all. Clonus is very rarely obtainable. It will be recalled that after spinal transection in lower animals, though flexor reflex activity soon returns after spinal transection, extensor reflexes, apart from the knee jerk which returns fairly promptly, are restored very slowly. "Spinal man" also shows a preponderance of the flexor reflexes. The flexor responses are, as mentioned above, well-developed before the knee jerk appears and become progressively more active, whereas, extensor responses, for the most part cannot be elicited. Stroking of the skin of the sole, for example, may result in a response involving a number of flexor muscles accompanied by inhibition of their antagonists. Sometimes contraction of the extensors of the opposite limb (crossed extensor reflex, p. 957) occurs. The hypertonicity of the anal and vesical sphincters becomes less pronounced. The bladder and rectum may empty automatically, that is, reflexly. When the stage of reflex activity is fully developed, spontaneous reflex spasms of the paralyzed limbs occur and quite a mild stimulus applied to the paralyzed limbs or to the genital region results in a widespread reflex contraction of flexor muscles. This response, which has been called the "*mass reflex*" by Head and Riddock, will be considered more in detail presently. Though the tone of all muscles is lower than normal, the flexors are less hypotonic than the extensors, and the limbs tend to be drawn into flexed positions. The paralyzed state is therefore referred to as a *paraplegia in flexion*. The duration of the stage of reflex activity is indefinite. The tracts whose destruction is essentially responsible for the features of paraplegia in flexion are the corticospinal and vestibulospinal. It will be remembered that after destruction of the ves-

tibular nuclei the extensor dominance of the decerebrate animal is abolished, whereas the flexor responses are retained and may even be increased.

III Stage of failure of reflex activity This stage is ushered in when, as a result of some infective complication (e.g., bed-sores, cystitis, pyelitis, etc.) and the consequent absorption of toxins, the spinal reflex centers are rendered functionless. As a rule it precedes death by a short interval. The extensor reflexes (e.g., knee jerks) are the first to disappear. Then the flexor responses are elicited with gradually increasing difficulty and the "mass reflex" does not occur. Finally all reflex activity is abolished and the muscles waste. Retention of urine and feces occurs, or there may be continual dribbling of urine, and fecal incontinence. Degeneration of the cells of the spinal gray matter, as a result simply of their isolation from higher centers (isolation dystrophy) which, according to Sherrington, occurs in the spinal monkey, does not apparently occur in spinal men for, as mentioned above, the stage of reflex activity may in the absence of infective conditions continue for indefinitely long periods. If toxemia commences in the first stage (spinal shock) this may merge into the third stage, the second stage of reflex recovery being absent.

INCOMPLETE DIVISION OF THE CORD

A bilateral lesion of gradual onset destroying the pyramidal tracts above the lumbar region but leaving the vestibulospinal fasciculi intact causes *paraplegia in extension*. The muscles are spastic and extensor activity predominates. The tendon jerks are exaggerated and patellar and ankle clonus can be readily elicited. Whereas paraplegia in flexion is analogous to the spinal state of animals, paraplegia in extension is more comparable to decerebrate extensor rigidity. The "extensor" plantar response, which is in reality a flexor protective reflex, is present, but much less intense and involves fewer muscles, than that seen in paraplegia in flexion. It is accompanied by a crossed extensor reflex. The mass reflex is absent. The abdominal reflexes are lost. Paraplegia in extension will also result from a bilateral lesion of the corticospinal tracts throughout any part of their course from the cortex downwards, and in a unilateral lesion (hemiplegia) the features are of the same nature. On the other hand, a lesion of the cord even though incomplete, if it involves the vestibulospinal tracts as well as the pyramidal, gives rise to the features characteristic of paraplegia in flexion.

The "mass reflex"

This, as already mentioned, occurs in conditions in which the spinal centers having been released from higher control establish independent activity and flexor responses predominate. A relatively weak stimulus causes, as a result of irradiation within the cord, a diffuse reflex response in which are included a large number of voluntary flexor muscles as well as the musculature of certain viscera. The response shows no "local sign," i.e., there is no circumscribed area from which it is elicitable alone, and it shows little or no modification upon varying the site of application of the stimulus. It can be evoked from anywhere over the limbs, genital regions or abdomen below the level of the lesion. The type of stimulus to which it responds is one which would be painful or unpleasant (e.g., scratching, pinching, pricking, etc.) could it be felt by the subject. It is a magnification of the normal withdrawal or protective reflex (flexion reflex, p. 957). The mass reflex comprises the following reactions:

(a) A *flexor spasm* of the muscles of the abdominal wall (recti) and lower extremities as a result of stimulation of the skin of the abdomen, genital region or limbs.

(b) *Sweating* caused by stimulation of the skin or of the vesical or rectal mucosae, its distribution is, roughly, that of the sensory loss.

(c) *Evacuation of the bladder*, even though only partially full, and of the *rectum* follows stimulation of the skin, or of the respective mucous surface. Evacuation of the bladder is affected normally by contraction of the detrusor muscle and relaxation of the sphincter (p. 483). Immediately after complete section of the cord the detrusor muscle may be capable of contracting reflexly as a result of the stimulus caused by distension of the bladder wall, but inhibition of the sphincter fails to occur, and retention of urine results. When, however, the sphincter is dilated by means of a catheter, the bladder wall contracts and evacuation occurs readily. After a time, relaxation of the sphincter, as well as contraction of the detrusor muscle, occurs through the reflex center in the sacral cord, when a sufficient quantity of urine (500 to 600 cc.) has collected to distend the bladder and so act as a stimulus. This reflex which characterizes the so-called *automatic bladder* and is brought about by intravesical stimulation is very readily facilitated by extra-vesical stimuli. Thus, as mentioned above, a scratch upon the sole of the foot or in the genital region causes a spasm of flexor muscles.

and, even though the quantity of urine present in the bladder at the time be quite inadequate to stimulate through distension alone, the extra-vesical stimulus has such a facilitating effect upon the vesical reflex that evacuation results. The facilitation caused by an extra-vesical stimulus is always more readily demonstrable upon the detrusor contraction (i.e., when a catheter is in position) than upon the sphincter. Though they have not been studied in the same detail the mechanisms concerned in the reflex evacuation of the rectum are no doubt similar.⁶

THE PLANTAR REFLEXES

This is the most appropriate place to consider these and other superficial reflexes.

The *normal plantar response* to a light scratch applied to the skin of the sole is plantar flexion of the four outer toes with no movement, or, more usually, plantar flexion of the great toe. The toe movements are accompanied by dorsiflexion of the ankle and contraction of the tensor fasciae femoris. The center for the reflex lies in the first sacral segment, its physiological significance is unknown. In a lesion of the corticospinal paths (see also p 1004) at any level above the first sacral segment, the normal response is replaced by one in which *dorsiflexion* of the great toe and often spreading or fanning of the outer toes occurs. It is elicited best from the outer border of the sole. This response is called after its discoverer the *sign of Babinski* (fig 66 15). From the dorsiflexion of the great toe, which is due to the contraction of the extensor longus hallucis, this reflex is also frequently referred to as the *extensor response*.⁷ This term, however, is incorrect since the upward movement of the great toe is part of a general flexor response homologous with the flexor reflex elicitable from the hind limb of a lower animal (p 957) (see Walshe). The dorsiflexors of the toes, though classed anatomically as extensors must,

when compared physiologically with similar muscles in the limb of an animal such as the dog, be included among the flexors. In lesions of descending spinal tracts (e.g., hemiplegia, paraplegia in extension) in which extensor reflex activity overshadows the flexor reactions, the so-called *extensor plantar response* is minimal, consisting mainly of the toe movements just described. The flexor nature of the reflex is shown, however, by the associated contraction of the ham-strings, i.e., semitendinosus, semimembranosus and biceps femoris, which invariably occurs (see also p 1012). Even when the toe movements do not occur, some slight contraction of the ham-strings can usually be detected. Its flexor nature is also indi-

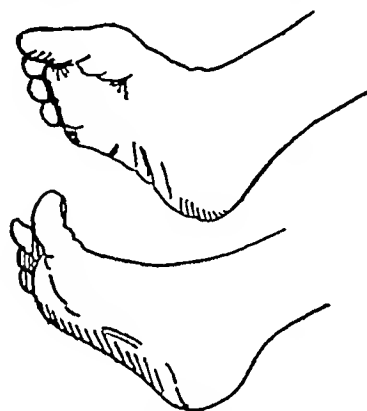


FIG 66 15 Upper drawing, normal plantar response, lower drawing, Babinski response

cated by the fact that an undoubted extensor reflex such as ankle clonus is readily inhibited by evoking the Babinski reaction (reciprocal inhibition). In hemiplegia in which the Babinski sign is present, stimulation of the sole of the *sound* side in some instances causes *plantar* flexion of the toes on the affected side together with contraction of the extensors of the knee and hip. This crossed response differs, however, from the normal plantar response (which of course is not crossed) in that contraction of the tensor fasciae femoris (a flexor) does not occur. It is looked upon as a *true* extensor response corresponding to the crossed extension reflex in the hind limb of the experimental animal when a flexion reflex is set up in the homolateral limb.

In the restricted forms of the extensor response as seen in hemiplegia, the receptive field of the reflex is also strictly circumscribed, being elicitable only from the sole—the outer border or the base of the hallux more especially.

In paraplegia in flexion in which flexor activity

⁶ Holmes has denied that the evacuation of the bladder in response to peripheral stimulation is an integral part of the mass reflex, i.e., due to irradiation of impulses to the micturition center in the cord, he attributes it simply to the sudden rise in intraabdominal pressure caused by the associated reflex contraction of the abdominal muscles.

⁷ An "extensor" plantar response is present normally in infants up to the first year or so, i.e., to the age of walking, and is then probably due to the undeveloped state of the corticospinal tracts. It is also present in normal adults during deep sleep (p 1066) and in the apneic stage of Cheyne-Stokes respiration, being apparently due in the latter instance, to anoxemia of the motor cortex.

predominates the so-called extensor response is maximal and, as we have seen, is simply part of a widespread flexor reaction. The receptive field of the reflex is extensive and the application of a stimulus is followed by contraction not only of the extensor longus hallucis (with dorsiflexion of the hallux) and ham-strings but by an associated contraction of the true flexors of the lower limb, the extensor longus digitorum, tibialis anticus, gracilis, sartorius, rectus femoris and iliopsoas. The extensors undergo reciprocal inhibition. The flexor nature of the "extensor" response is therefore undoubted. It is a nociceptive reflex, the limb, when the reflex is fully developed, being withdrawn from the stimulating (nocuous) agent by a flexion at hip and knee and dorsiflexion at the ankle.

OTHER SUPERFICIAL REFLEXES

(a) *Oppenheim's reflex* is simply a modified Babinski response, it is associated with the same conditions as the latter and has a similar significance. It consists in dorsiflexion of the hallux which results when a firm downward sliding pressure is applied to the skin over the postero-internal border of the tibia.

(b) "*Fanning*" of the toes, originally described by Babinski, occurs in lesions of the premotor area or of extrapyramidal fibers. It consists of lateral deviation of the toes when the sole is stroked.

(c) *Nociceptive reflexes of upper limb*. A flexor reflex of the upper limb corresponding to the sign of Babinski of the lower limb is often elicitable in lesions involving the pyramidal tracts above the thoracic region, as in hemiplegia. It consists of flexion of the fingers, often accompanied by flexion of wrist and elbow and abduction and external rotation of the shoulder, when a mildly nocuous stimulus is applied to the palm or

surfaces of the finger. Sometimes an extensor response—elevation, adduction and internal rotation of the shoulder, extension of the elbow, pronation of the forearm, flexion of wrist and hyperextension and adduction of the fingers—may be elicited in a pyramidal tract lesion by stimulation of the skin of the axilla or side of the chest.

(d) *Abdominal reflexes*. Lightly scratching the skin of the abdomen of a normal person causes a reflex contraction of the abdominal muscles. The reflex is abolished in pyramidal lesions or in one (e.g., acute anterior poliomyelitis) involving the centers located in the 7th to 12th thoracic segments. They do not appear until between the 6th and 8th month of age when the infant can sit up unsupported.

(e) *Cremasteric reflex* consists of contraction of the cremaster muscle and elevation of testicle which results from a light stroke applied to the skin on the inner aspect of the upper part of the thigh. It is abolished in pyramidal lesions or as a result of destruction of the center in the 1st or 2nd lumbar segment.

(f) *Bulbocavernosus reflex* has its center in the 3rd and 4th sacral segments. It consists of contraction of the bulbocavernosus muscle (detected by palpation) in response to stimulation of the glans penis. It is absent in a lesion involving any part of its reflex arc, motor or sensory limb, or center.

(g) *Anal reflex* is the contraction of the external anal sphincter in response to scratching the neighboring skin. Its center is situated in the 4th and 5th sacral segments and the coccygeal segment, it is lost after interruption of its reflex arc.

(h) *Gluteal reflex*. Scratching the skin of the buttock causes contraction of the gluteal muscles. It depends upon the integrity of the 4th and 5th lumbar segments.

The *deep or tendon reflexes*, such as the knee jerk, ankle jerk, etc., have been discussed in chapter 65.

THE EXTRAPYRAMIDAL SYSTEM THE THALAMUS AND HYPOTHALAMUS

There are at least two pathways other than the pyramidal or corticospinal through which the cerebral cortex exerts an influence upon the activity of the skeletal muscles, namely through the pons and cerebellum—the *corticopontocerebellar*—and through the striate body and the subthalamic nuclei—the *corticostriomgral*. The first mentioned of these two systems will be considered

and will be considered here. Though this system is associated functionally with the pyramidal system, and its fibers as well as those of the latter system pass through the internal capsule to reach their subcortical connections, they take a different course to the spinal cord, traversing the reticular formation of the brain stem, and the reticulospinal tracts. The extrapyramidal system is much more

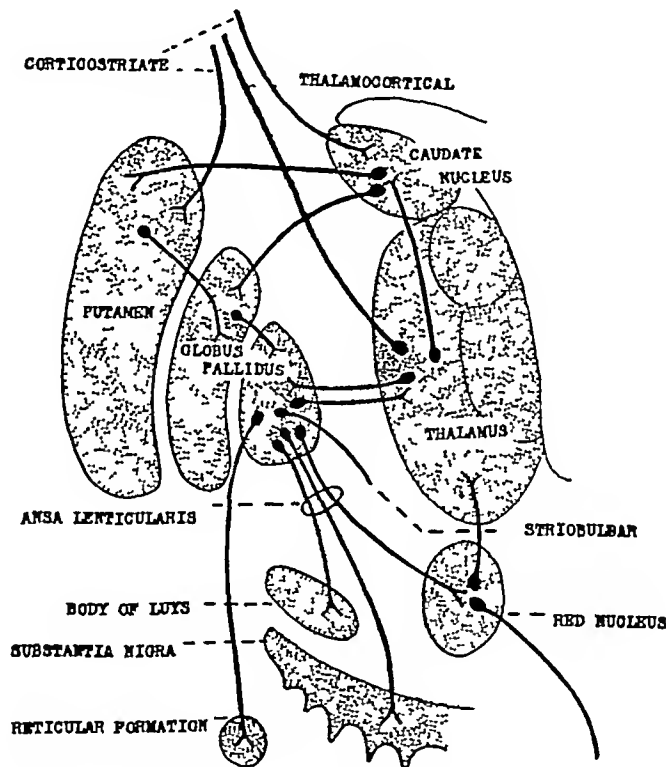


FIG 67 1 The corpus striatum and its connections, claustrum and external capsule not shown

in chapter 70. To the corticostriomgral system, together with the paths to the spinal cord, the term “extrapyramidal” is most usually restricted,¹

¹Strictly speaking, the term “extrapyramidal” should include all descending paths, other than pyramidal, that are concerned with the motor functions, and some authors do so. A. Brodal, for example, recognizing the difficulties of definition, states, “From a practical point of view it appears most appropriate to regard the extra-pyramidal system as consisting of all tracts and nuclei which form the substrate for motor functions, other than the pyramidal tract and its site of origin in the cerebral cortex.” This definition would embrace the corticopontocerebellar tracts

extensive than the pyramidal. The highest level of this system is situated mainly in the premotor area of the cerebral cortex. Impulses which travel by extrapyramidal pathways can be evoked from parts of the frontal cortex lying in front of the motor area (area 4). The responses which result are more complex and general in nature than those initiated from the motor area itself (ch 68). They consist of movements associated with postural adjustments, turning of the eyes and head, movements of the limbs and twisting of the trunk. Such movements can still be elicited after excision

of the motor area, but not after projections from the premotor area (area 6) are interrupted by undercutting (Figs 67.1 and 67.2)

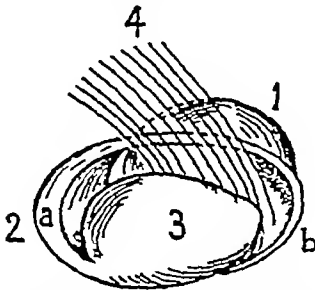


FIG 67.2 The corpus striatum 1, thalamus, 2, caudate nucleus, a, head, b, tail, 3, lentiform nucleus, 4, fibers of internal capsule

lying on the medial side of the anterior limb of the internal capsule. Its head bulges medially into the anterior horn of the lateral ventricle. The tail ends in the *amygdaloid nucleus*, a small globular mass of gray matter whose cells constitute a part of the olfactory pathway. The *lentiform nucleus* occupies a corresponding position on the outer side of the internal capsule (fig 66.12, ch 66). The wedge-shaped lentiform nucleus is divided again into an outer and an inner part, the *putamen* and the *globus pallidus*, respectively. Beneath the anterior limb of the internal capsule the lentiform nucleus becomes continuous with the head of the caudate nucleus. On the outer side of the lentiform nucleus is a narrow band of white matter—the *external capsule*, lateral to this again is an elongated island

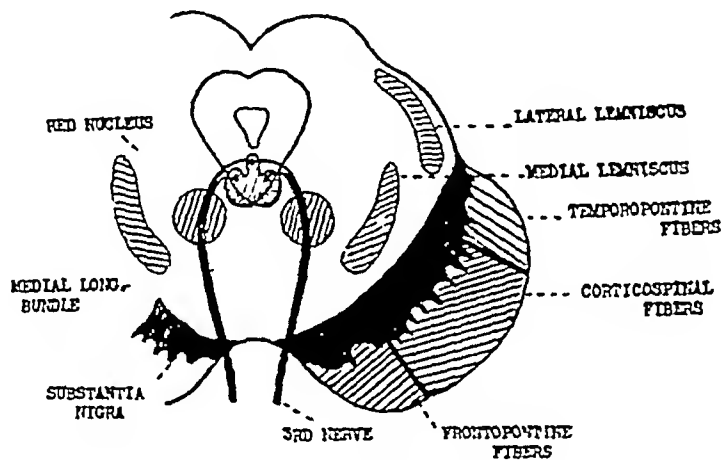


FIG 67.3 Diagram of section through the mid brain at the level of the red nucleus

The corticostriatonigral system embraces four masses of gray matter, (a) the *corpus striatum* (basal ganglia) (b) the *red nucleus*, (c) the *substantia nigra* and (d) the *body of Luys* (subthalamic nucleus). These several nuclear masses are connected together by numerous fiber tracts, and give rise in turn to descending fibers through which connections are made with the spinal motoneurons.

THE CORPUS STRIATUM

The corpus striatum is a mass of gray matter lying at the base of the hemisphere in close relation to the thalamus. It is incompletely divided into two main parts—the caudate and lentiform nuclei—by the anterior limb of the internal capsule. The *caudate nucleus* is a gray mass with a large pear-shaped head and a long curved tail

of gray matter—the *claustrum* (fig 67.2). The corpus striatum, amygdaloid nucleus and claustrum are together known as the *basal ganglia* (or *nuclei*).

The fiber system of the corpus striatum. The fiber connections may be divided into the following groups

(1) *Fibers arising and ending within the corpus striatum (internuncial).* These pass from (a) the putamen to the globus pallidus (b) from the caudate nucleus to the putamen and globus pallidus, and (c) from the lateral to the medial zone of the globus pallidus (fig 67.1).

(2) *Fibers arising in the corpus striatum and ending elsewhere (striofugal fibers).* The great bulk of the fibers leaving the corpus striatum arise from the globus pallidus. (a) *Striothalamic* from globus pallidus to the

lateral nucleus of the thalamus. Through these, and the thalamo-olivary and olivospinal tracts a possible pathway exists for striatal impulses to reach the spinal cord. (b) *Striosubthalamic (ansa lenticularis)*. These arise in the globus pallidus and pass to the red nucleus, substantia nigra and the body of Luys. (c) There is no direct connection, apparently, between the corpus striatum and the spinal centers though, according to Morgan, a direct path (striobulbar) to the bulbar nuclei exists. These fibers arise in the globus pallidus. They terminate around cells in the reticular formation of the pons and medulla and in the trochlear, abducens, trigeminal (motor), glossopharyngeal and hypoglossal nuclei, of the same and the opposite side. The reticulospinal and rubrospinal tracts provide pathways whereby striatal impulses reach the spinal centers. (d) *Pallidohypothalamic* from the globus pallidus to the hypothalamus. No fibers passing from the striate body to the cortex were found by Wilson, who made a comprehensive study of the striatal connections. Others, however, state that such exist.

(3) *Fibers ending in the corpus striatum but arising elsewhere (striopetal fibers)*. (a) *Thalamus to caudate nucleus, putamen and globus pallidus*. (b) *Body of Luys to globus pallidus*. (c) *Red nucleus to globus pallidus*. (d) *Cortex to caudate nucleus*. Area 6 (premotor) sends fibers to this part of the corpus striatum and possibly also to the putamen and globus pallidus.

(4) *Fibers joining the corpora striata of the two sides (commissural fibers)*. These pass from the globus pallidus of one side to the corresponding nucleus of the opposite side.

(5) *Fibers passing through the corpus striatum (fibers of passage)*. Ascending and descending fibers of internal capsule—corticospinal, thalamocortical, etc.

The corpus striatum is also almost certainly in communication with the cerebellum through relay stations, e.g., red nucleus.

THE RED NUCLEUS

The red nucleus is a large oval mass of gray matter in the tegmentum of the mid-brain lying beneath the thalamus and extending backwards from about the middle of the latter structure to a level beneath the posterior border of the superior colliculus (figs 671 and 673). The fibers of the third nerve stream through it. Two groups of cells are found within it. The cells of one of these groups are large (*nucleus magnocellularis*) and are situated at the posterior part of the red nucleus. The other, situated anteriorly, is composed of small cells (*nucleus parvocellularis*). Phylogenetically, the nucleus parvocellularis is a later acquisition. Fibers arising from the nucleus magnocellularis of one side issue from the ventral aspect of the red nucleus and, crossing with those of the opposite

side (Forel's decussation) descend as the *rubrobulbar* and *rubrospinal tracts*. Fibers descend from the small-celled nucleus to synapse with cells in the reticular formation of the pons and medulla, they constitute the rubroreticular tract. The following is a summary of the connections of the red nucleus.

A Efferent tracts (rubrofugal)

(a) Rubrobulbar and rubrospinal tracts from nucleus magnocellularis to the motor cells of medulla and cord.

(b) To the reticular formation of the pons and medulla (tractus rubroreticularis).

(c) Rubro-olivary tract, to the inferior olivary nucleus.

(d) Rubrothalamic tract to the lateral nucleus of the thalamus.

(e) Rubrostriatal to globus pallidus.

B Afferent tracts (rubropetal)

(a) Corticorubral tract, from the frontal lobe and premotor area to the nucleus parvocellularis.

(b) Pallidorubral tract from the globus pallidus.

(c) Cerebellorubral tract (p. 1077) composed of fibers which arise mainly in the dentate nucleus of the opposite cerebellar hemisphere and reach the red nucleus via the superior cerebellar peduncle. This cerebellar pathway is relayed to bulbar and spinal centers by the rubrobulbar and rubrospinal tracts, and to the thalamus for relay to the cerebral cortex.

Little is known definitely with regard to the functions of the red nucleus. It appears, however, to be an essential part of the extrapyramidal pathway controlling the performance of complex muscular movements. It may be regarded as a center wherein impulses received from various sources (cerebral cortex, corpus striatum, cerebellum, etc.) are organized before their transmission (via rubrobulbar, rubrospinal and rubroreticular tracts) to the motor centers of the cranial and spinal nerves.

THE SUBSTANTIA NIGRA

The substantia nigra is the crescentic gray mass seen in transverse section lying ventral to the red nucleus and between the basal and tegmental portion of the cerebral peduncles (figs 671 and 673). It is composed of large cells which are deeply pigmented with melanin and have a high content of iron. It is connected by both afferent and efferent fibers with the globus pallidus and the frontal region of the cerebral cortex. It also receives fibers from the body of Luys (subthalamic nucleus) and from the medial and lateral lemnisci, the superior colliculus and the mam-

millary bodies. It is thus in receipt of afferent impulses from the general body surface and from the organs of hearing, sight and smell. It sends fibers to the red nucleus and to the formatio reticularis of the pons. The substantia nigra is regarded as a center for the integration of those afferent impulses which are essential for the performance of skilled movements. It shows its greatest development in man and the higher apes.

THE BODY OF LUIS OR SUBTHALAMIC NUCLEUS

The anterior limit of this structure lies on a level with the mammillary bodies. It should not be confused with the hypothalamic nuclei (p 1026). It lies lateral and ventral to the red nucleus and dorsal to the substantia nigra. Fibers link it to both these structures and to its fellow of the opposite side. It receives fibers from, and sends a few fibers to the globus pallidus (p 1016).

THE EFFECTS OF EXPERIMENTAL STIMULATION AND INJURY OF THE CORPUS STRIATUM

Phasic movements in any way resembling those which can be elicited by excitation of the motor cortex cannot be induced by stimulation of the basal ganglia or of the subthalamic region. When such movements occur they are attributed to the escape of current to the internal capsule. A slow movement of the legs of cats was observed by Miller upon unipolar faradization of the caudate nucleus or by the application of warmth or of strychnine, which appears to be due to excitation of this nucleus itself. But outstanding result of stimulation of the striate body is observed during muscular movements initiated from the cortex or when the limbs are held by tonic postural contractions, then inhibition promptly occurs. Among the most definite experimental results are those of Mettler and his associates. They observed the following effects of electrical stimulation of the corpus striatum in monkeys: (a) Movements induced by excitation of the motor cortex were inhibited by concurrent stimulation of the caudate nucleus, putamen or claustrum, (b) stimulation of the globus pallidus added to movements initiated from the motor cortex, an element of "plastic tonus" which exerted a "holding" action upon the movements and prolonged their reaction time.

Experimental lesions of the striatal region in dogs or cats result in no noticeable motor defect, nor in monkeys is any pronounced effect observed. Kinnier Wilson found that when extensive uni-

lateral lesions were made the animals used by preference the arm and hand of the ipsilateral side and showed clumsiness of movement on the opposite side. There was, however, no paralysis or even paresis. In chimpanzees, striatal injury is followed by athetoid movements (p 1020) closely similar to those seen in persons suffering from disease of this region. Brown observed tremors in monkeys following large bilateral lesions in the putamen or smaller bilateral lesions of the globus pallidus. Liddell and Phillips observed slight but persistent hypertonus in the extensor muscles of the opposite side and flexor hypertonus on the same side following electrolytic lesions in the corpus striatum. Placing reactions were imperfect and closure of the contralateral eyelid was defective. The highest integrative level of the corticostriatonigral system consists of cells situated chiefly in the premotor area (area 6) (ch 68).

FUNCTIONS OF THE EXTRAPYRAMIDAL SYSTEM

The corpus striatum is one of the oldest parts of the cerebrum, and the globus pallidus (*paleostriatum*) is older phylogenetically than the putamencaudate portion (*neostriatum*). The connections of the paleostriatum are chiefly efferent, the neostriatum is mainly receptive. In lower vertebrates (e.g., fishes, amphibia, reptiles and birds), in which the cortical mantle is absent or rudimentary and the pyramidal system has not yet come into being, the corpus striatum² is the highest motor center, being looked upon as homologous with the motor cortex of higher forms. Upon this "old motor center" and its connections with lower levels the motor activities of sub-mammalian forms very largely depend. In the bird, for instance, after removal of its rudimentary cortex, visual and auditory sensations seem unaffected and it continues to carry out normal movements of feeding, courting, fighting, etc. These instinctive reactions are imperfectly executed after removal of portions of the striate body. As the phylogenetic development of the cerebral cortex and the pyramidal tracts progresses, corticalization of function proceeds accordingly. But the functions of the extrapyramidal system, though in some ways subordinated to those of the motor cortex, are by no means lost. The phylogenetically older system continues to

² The infant's movements are of such a nature as to suggest that they are under the control of the extrapyramidal system, the pyramidal system being as yet not fully developed.

carry out those more or less automatic or semi-reflex movements concerned with the maintenance of posture, defence, feeding, etc. In the cat or dog for example decortication is followed by relatively little disturbance of the motor functions as compared with the effects of this operation in primates. The motor disability of the decorticated monkey is profound, whereas the dog a short time after decortication is able to walk about, and responds to stimulation by growling or barking, and can feed itself, even choosing appetizing from disagreeable food. The extrapyramidal paths undoubtedly play an important rôle in the ordinary motor activities of these animals. In man the cortical representation of the extrapyramidal system is placed in the premotor area (p 1033). From the results of ablation experiments involving the premotor cortex (p 1034), this frontal region in primates appears to have become intimately associated, through intracortical connections with the motor area, serving to synthesise larger complex actions from the smaller more discrete movements governed by the latter area. Wilson's view of the functions of the extrapyramidal striatal system in the human subject is that through the convergence of its paths upon the spinal motoneurons it maintains a postural background or substructure, against or upon which voluntary movements are executed. Impulses traveling these paths exert, he believes, a steadying effect upon, but do not themselves initiate, such movements. Kappers, however, considers that the habitual or automatic acts of daily life are mediated through this extrapyramidal system and points out that such types of movement suffer the greatest impairment in striatal disease.

CLINICAL MANIFESTATIONS OF DISEASE OF THE EXTRAPYRAMIDAL SYSTEM

The chief clinical features of extrapyramidal disease are (a) muscular rigidity resulting in disturbances of posture and movement, (b) involuntary movements, e.g., tremor, athetosis, chorea, (c) absence of any true paralysis. The corpus striatum and other parts of the extrapyramidal system seem to be peculiarly susceptible to the actions of certain toxic substances. The following are some of the syndromes met with (1) *Progressive hepato-lenticular degeneration*, (2) *Parkinsonism*—paralysis agitans, etc., (3) *chorea*, (4) *athetosis*, (5) *tortion spasm*.

Progressive hepato-lenticular degeneration (Wil-

son's disease) This was described by Wilson in 1912. It invariably terminates fatally but its duration varies from a few months to several years. The following are its chief features.

(a) *Muscular rigidity* is wide-spread and progressive, it involves face, trunk and limbs. Flexors as well as extensors are affected, but the former more conspicuously than the latter. The hypertonus offers a "lead-pipe-like" resistance to passive movement and results in slowness and difficulty of movement. Eventually, contractures occur and the patient is rendered almost immobile as though carved from stone, he can be lifted or moved *en bloc*. The rigidity of the facial muscles gives a fixed, blank expression. The mouth is sometimes held widely open, the smile or laugh is peculiarly stiff and vacuous. The hypertonus of the muscles of articulation and deglutition leads to dysarthria (p 1049) and dysphagia. The rigidity is temporarily abolished by the injection of novocaine into the muscles.

(b) *Involuntary movements*. These consist chiefly of tremor (about 6 oscillations per sec.) which is increased by excitement or any attempt at voluntary movement, sometimes athetoid movements (p 1020) occur.

(c) *The reflexes are normal*. There are no sensory changes and although the muscles often show some weakness and are easily fatigued there is no actual paralysis.

(d) *Cirrhosis (multilobular) of the liver* is found at autopsy, but during life there may be no signs of liver disease. In some instances, however, symptoms pointing to the liver precede the nervous manifestations.

(e) *Emotionalism*. Involuntary laughing or crying, and some mental deterioration.

(f) *Greenish brown pigmentation of the cornea (in Descemet's membrane)* occurs in a proportion of cases.

Degeneration of the cells of the putamen and sometimes cavity formation are found at autopsy. The caudate nucleus and globus pallidus are affected to a much less degree. A toxin of some sort is probably responsible for the disease. It is possible that the toxic substance is absorbed from the alimentary canal, and damages both hepatic and nervous tissue.

The Parkinsonian syndrome. The principal features of this syndrome are the following. (a) A *coarse tremor* involving head and limbs. The hand may show "pill-rolling" movements, i.e., rhythmic movements of thumb upon the first two fingers, alternating movements of flexion and extension at the wrist, or of supination and pronation of the forearm, are frequently present. When the limb is engaged in some voluntary act the movements often temporarily disappear in that limb but become more pronounced in other parts. The tremor becomes more pronounced during

emotion Since it affects those muscles concerned mainly with the maintenance of posture and disappears during sleep, it is not actually a tremor of rest as is sometimes stated It disappears on the paralysed side if hemiplegia supervenes, ablation of the motor area of the cortex, or section of the lateral corticospinal tracts also abolishes it in the contralateral side It is therefore dependent upon an intact pyramidal tract. (b) *Muscular rigidity* which leads to slowness and stiffness of movement and a fixed mask-like facies (Parkinson's mask), emotional expression is "wiped" from the face The patient winks infrequently, and speech is slow The rigidity is different both in its quality and distribution from that characteristic of the decerebrate animal or of the hemiplegic patient (p 1004), nor has it the even, smooth, plastic (lead pipe) quality seen in hepato-lenticular degeneration Flexors and extensor muscles are affected about equally, and the resistance of a limb to passive movements has been described as resembling cogwheels moving slowly upon one another, as though groups of muscles gave way in succession to the stretching force The upper limbs are held in characteristic attitudes of adduction at the shoulders, flexion at the elbows, flexion or slight extension at the wrists, flexion at the metacarpophalangeal joints and slight flexion at the interphalangeal joints (c) The *gait* is slow and shuffling with short steps, or it may be "festinating" in character, i e., the patient is bent forward and hastens along with short quick steps as though trying to "catch up to his center of gravity" and prevent his falling When pushed forward or backward he cannot stop quickly but moves by a series of small rapidly repeated steps in the direction in which he is pushed *Propulsion* and *retropulsion* are the respective terms applied to these forward and backward movements There is no true paralysis, the reflexes and sensation are unaffected The condition is due to degenerative changes in the corpus striatum, according to Hunt, the chief lesion is a progressive atrophy of the globus pallidus Degeneration of the muscle spindles also occurs

The picture of Parkinsonism just drawn is one which is seen in elderly persons, it occurs without apparent cause and is slowly progressive It is also spoken of as *paralysis agitans*, and is probably due to a senile vascular (arteriosclerotic) change in the striate nuclei Other forms of Parkinsonism, essentially the same in their manifestations but differing in detail may occur at any age and

develop more rapidly as a result of *epidemic encephalitis* (encephalitis lethargica), and are due apparently to the action of the virus of the disease upon the striate body Rigidity is the most prominent feature of the postencephalitic type—choreic and athetoid movements may also occur The patient may assume statuesque attitudes for long periods—*catalepsy*—or maintain the limb in a position which has been passively imposed upon it. The substantia nigra is particularly chosen for attack in this type The locus coeruleus in the floor of the 4th ventricle is also affected in a proportion of cases, and this fact may account for the disturbances in the autonomic nervous system which is sometimes a prominent feature (McAlpine)

Chorea There are two principal forms of this condition

(1) *SYDENHAM'S CHOREA* (or St. Vitus's dance) is not uncommonly a sequel to acute rheumatism Its chief features are (a) involuntary jerky movements, semi purposeful in character, involving the muscles of the limbs and face They are intensified by excitement but disappear during sleep The mobility of the face is in marked contrast to the fixed "starched" expression of Parkinsonism There may be athetoid movements, the condition then being termed *atheto-chorea* (b) *Hypotonia* of the muscles

Sydenham's chorea may be bilateral or unilateral (hemichorea) Death is rare and there is consequently uncertainty concerning its neuropathology The lesions are probably similar in nature to those responsible for athetosis Degenerative changes in the corpus striatum (chiefly in the putamen and caudate nucleus), substantia nigra and body of Luys have been described in fatal cases Hemorrhage into the body of Luys of one side has been found in hemichorea, or it may result from an irritative lesion of the premotor area

(2) *HUNTINGTON'S CHOREA* is a much more serious affection, being usually fatal It is hereditary, showing Mendelian dominance Of about 1000 cases arising in the United States practically all can be traced to some half dozen individuals, including three brothers who settled there in the 17th century The choreiform movements are often violent, there are dysarthria, ataxia of the limbs and progressive dementia The changes in the central nervous system are, marked atrophy of the cerebral cortex and the corpus striatum Of the latter, the putamen and caudate nucleus are involved, the globus pallidus escapes

Athetosis (mobile spasm) This is the term applied to involuntary movements of a peculiar slow writhing, twisting or squirming character involving the upper limbs and less commonly the face and lower limbs They may be unilateral or bilateral They are increased when any voluntary movement is attempted but dis-

appear during sleep. The muscles when not actually engaged in the abnormal movements are hypotonic. As seen in the hand the movements consist of alternate extension and flexion at the wrist and metacarpophalangeal joints, with the fingers usually held extended at the interphalangeal articulations. Involvement of the facial muscles results in grotesque facial expressions (grimacing), involvement of the muscles of the mouth and throat causes disturbances of articulation and deglutition.

Not uncommonly the involuntary movements are both choreiform and athetoid in character, when the term *choreo-athetosis* is applied.

The neural basis of these conditions is not known precisely, but the following observations point very definitely to the implication of the extrapyramidal system.

(a) In these diseases lesions have been reported in the premotor area, corpus striatum, substantia nigra, subthalamic nucleus, and in the thalamus, connections of the latter with the striate body being especially involved. The striatal structures may show a characteristic appearance known as *status marmoratus* or *état marbré*, in which the tissue on cross section is streaked or blotched like marble.

(b) Tremor, athetoid movements and rigidity have been produced experimentally especially by large bilateral ablations of the caudate nucleus and putamen, but smaller excisions are effective if area 6 is also removed.

(c) Bucy and Buchanan have reported the case of a child whose athetoid movements were abolished permanently by the excision of the central part of the precentral gyrus (i.e., area 6 of the premotor cortex, p. 1032) leaving the motor area almost intact.

Bucy also found that barbiturates which depress area 6 (premotor area) but have little effect upon the projections of the motor area, temporarily abolish athetosis. He believes that the abnormal movements are due to release from suppressor influence (p. 1039) of extrapyramidal projections from the premotor area (see below).

(d) Division of the anterior columns of the spinal cord (which transmit the reticulospinal tracts) without injury to pyramidal fibers abolishes the athetoid movements (Putnam).

Bucy proposes the theory that the choreo-athetoid movements are due to the release of extrapyramidal activity of an abnormal nature in area 6 and 4 by the interruption of a neural circuit which normally exerts a suppressor function (ch. 68) upon these areas. This circuit, starting in areas 8 and 4, descends to the caudate nucleus, passes to the globus pallidus, thence through the ansa lenticularis to the midbrain, then ascends to the anterolateral part of the ventrolateral nucleus of the thalamus, and finally (via thalamocortical fibers) returns to the cortex, ending mainly in area 6, but also in area 4.

Torsion spasm is a very rare condition and need only be defined. It consists of spasm of neck, trunk and limb muscles which twist the body into bizarre attitudes. The muscles, following the spasm, are hypotonic. Pathological changes in various parts of the extrapyramidal system have been described.

THE THALAMUS

This large gray mass is related medially to the 3rd ventricle which lies between the thalami of the two sides. The thalami are joined across the midline by an isthmus, the *massa intermedia*. The posterior limb of the internal capsule lies upon the outer side of the thalamus and separates it from the lentiform nucleus. Above the thalamus is the lateral ventricle, a part of whose floor it forms. In front is the head of the caudate nucleus, the arched *body* of the latter is related to the upper part of the lateral surface of the thalamus. Below the thalamus are the corpus of Luys (subthalamic nucleus) and the forepart of the red nucleus (figs. 67 1 and 67 3).

There are five main nuclear masses in the thalamus.

A vertical septum of fibers known as the *internal medullary lamina* divides the principal part of the thalamus into A, a *medial* and B, a *lateral* mass, each of which contains two main nuclear groups. C. In the *massa intermedia* and the adjacent part of the medial mass are discrete groups of nerve cells known as the *nuclei of the midline*. D. Clusters of nerve cells are present in the internal medullary lamina itself—the *intralaminar nuclei*. E. *Pulvinar*.

A. *Nuclei of the medial mass*. 1. *Anterior nuclei*. These form a mass which bulges into the lateral ventricle. They receive fibers from the *mammillary bodies* (mamillothalamic tract of Vicq d'Azyr) which convey olfactory impulses. They send fibers to the *paracentral lobule* and the *posterior part* of the *cingular gyrus* on the medial aspect of hemisphere. 2. *Dorsomedial nuclei*. These consist of a dorsolateral group of small cells, and a medial collection of large cells. The former projects to prefrontal areas of the cerebral cortex. The large-celled portion is connected by both afferent and efferent fibers with the hypothalamus, it also projects to the corpus striatum, and has rich afferent connections with other thalamic nuclei. The dorsomedial nuclei are thought to serve as an *association center* where visceral and crude somatic sensations are synthesized. It is thought to be a conscious center for the cruder (protopathic)

sensations, and where sensations are integrated into "feelings", both pleasant and unpleasant

B Nuclei of the lateral mass *I. Ventral group* This consists of, (a) *Anterior ventral nucleus* which receives fibers from the globus pallidus and projects to different parts of the corpus striatum, but not to the cerebral cortex. This nucleus occupies the most anterior (rostral) extremity of the lateral mass (b) *Lateral ventral nucleus* The fibers entering the posterior and medial part of this nucleus are derived mainly from the dentate nucleus of the contralateral half of the cerebellum (via the dentatothalamic and the dentatorubrothalamic tracts) Its efferent fibers pass to area 4, and in

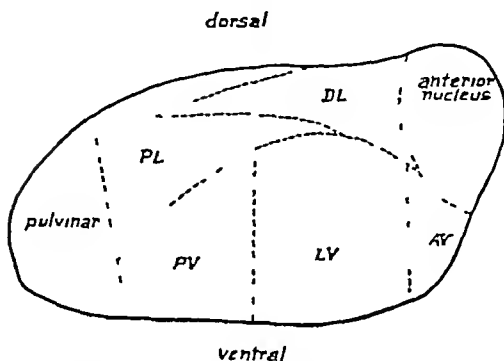


FIG 67.4 Showing the anterior nucleus and the divisions of the lateral nuclear mass of the thalamus DL, dorsal lateral nucleus, PL, posterior lateral nucleus, PV, posterior ventral nucleus, LV, lateral nucleus, AV, anterior ventral nucleus (kindness of Dr Murray Barr)

much smaller numbers to area 6 of the cerebral cortex. Thus a pathway is established through which voluntary movements can be brought under the influence of the cerebellum (c) The *posterior ventral nucleus* is the main subcortical center for sensory impulses ascending in the trigeminal, medial and spinal lemnisci (proprioceptive from muscles and joints, light touch, discrimination of two points, pain, heat and cold) It projects to the post central gyrus (somesthetic area), areas 1, 3, 2, 5 and 7. This nucleus also sends fibers to the hypothalamus and corpus striatum. The posterior ventral nucleus is composed of two subsidiary parts: (i) the *posterolateral nucleus*, which receives the trigeminal fibers, and (ii) the *posteromedial nucleus* which is the thalamic station for the medial and the spinal lemnisci.

II Dorsal group This is also called the *lateral nucleus of the thalamus*. It is subdivided into an anterior and a posterior portion, designated,

respectively, the *dorsolateral* and *posterolateral nuclei*.

C The nuclei of the midline are, as mentioned above, situated in the massa intermedia, and in the adjacent part of the medial mass forming the upper part of the wall of the third ventricle. They are phylogenetically the oldest of the thalamic nuclei and are a center for the most primitive

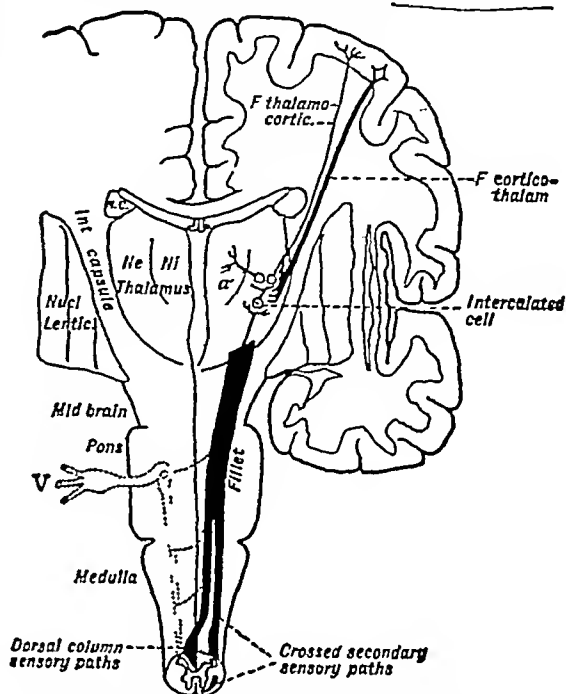


FIG 67.5 Diagram showing the paths and centers concerned in sensation. All sensory impulses ascend to the lateral part of the thalamus, where regrouping occurs, the cruder sensations (e.g., those of pain and extremes of temperature) are relayed to the medial portion of the thalamus, the remainder (e.g., light touch, discrimination of two points, sense of position and movement, etc.) to the cerebral cortex. The cortico-thalamic fibers, which terminate in the lateral nucleus of the thalamus are also shown (modified from Head)

forms of sensation, e.g., from the viscera and other structures occupying the axial regions of the body. The mid-line nuclei are connected by many fine myelinated fibers with other subcortical centers, namely, the hypothalamus and mid-brain nuclei. They also receive fibers from the corpus striatum. These nuclei have many intrathalamic connections but few if any fibers pass to the cerebral cortex, corpus of Luys (subthalamic nucleus) and the forepart of the red nucleus.

D The intralaminar nuclei are scattered groups of cells within the internal medullary lamina. The

connections of these nuclei are imperfectly known, but they are thought to receive both trigeminal fibers and fibers of the medial lemniscus, they project to the globus pallidus and have many connections with neighboring thalamic nuclei. A well-defined nucleus in the middle of the thalamus, and usually classed with this group is known as the central or centromedian nucleus. It is especially well-marked in primates, as is possibly an intra-thalamic integrating center.

E The pulvinar is the expanded continuation posteriorly of the lateral nuclear mass. It overhangs the superior colliculus. It projects to the cortex, its inferior part to an area comprising the posterior part of the temporal lobe, and the anterior part of the occipital lobe. The rest of the pulvinar projects mainly to the posterior parts of the parietal and temporal lobes.

The medial and lateral geniculate bodies are usually included as part of the thalamus or referred to as the metathalamus, they are dealt with in chapters 76 and 77.

It will be seen from the foregoing account that the nuclei of the thalamus can be divided upon a functional basis into three categories, namely, (a) those which serve as relay stations of afferent impulses from the periphery to the cortex, (b) those which are connected mainly with other subcortical centers and (c) those whose chief function is associative.

Functions

We have seen that the corpus striatum is a part of an old or primitive motor system. The thalamus, on the other hand, is a primitive receptive center wherein sensory impulses give rise to a *crude uncritical form of consciousness*, a consciousness which is greatly elaborated upon, especially in man, by the cerebral cortex (see ch 68). The thalamus serves as a great integrating center, through its connections with the corpus striatum, tactile, painful, olfactory and gustatory impulses are correlated with motor reactions. Sensory localization in the thalamus has been demonstrated by Dusser de Barenne and Sager by means of local strychninization. The injection of a minute quantity of the drug into the thalamus in cats is followed by hyperesthesia and hyperalgesia to cutaneous and deep (muscle, tendon and periosteum) stimuli. The cutaneous hypersensitivity is most pronounced on the contralateral side of the body, increased sensitivity to deep stimuli occurs only on the contralateral side. By this method,

sensory areas in the posterior ventral nucleus of the thalamus for the head, arm and leg have been located. These areas are not, however, sharply delineated as in the cerebral cortex but show rather wide overlapping.

The thalamus is not, however, entirely sensory in function. In animals possessed of little or no cortex, or in higher mammals (e.g., cat or dog) after decortication, it and the corpus striatum serve for the execution of complex movements of an automatic or reflex nature. Motor responses also result from its direct stimulation. Conjugate deviation of the head and eyes to the opposite side and movements of progression in the limbs follow electrical excitation of the ventral nucleus. The pulvinar has in the past been looked upon as a relay station in the transmission of visual impulses to the occipital cortex. It now appears, however, from the researches of Brouwer and Zeeman that though this part of the thalamus receives a few fibers of the optic tract, it is not a primary visual center (see p 1171).

The regrouping of afferent impulses within the thalamus

The clinical researches of Head and Holmes indicate that the various types of sensory impulses which reach the thalamus are regrouped. The cruder, more essentially primitive sensations (protopathic) e.g., *pain, extremes of temperature, pressure, contact over hair-clad parts* terminate in the thalamus itself, probably in one or other of the medial groups of nuclei, after being relayed from the posterior ventral nucleus. The impulses received by this part of the thalamus are thought to arouse a relatively vague type of consciousness or "crude awareness"—sensations which come under the category of "feelings" (affective sensations), both pleasurable and disagreeable. They are devoid of discriminative qualities. Fibers subserving the finer discriminative sensations are relayed to the somesthetic area of the cortex and ascend through the internal capsule to give rise to the following sensations: *light touch over hairless parts, localization, two point discrimination, the appreciation of gradations of temperature* (between 20° and 40°C.) and the *sense of position and movement* of parts of the body. Unconscious proprioceptive impulses from muscles and joints are also relayed from the cerebellum through the thalamus. In the chimpanzee, bilateral decortication permanently abolishes the sense of position and movement, but pain can still be appreciated.

Kinesthetic sense is only contralaterally represented in the thalamus, but touch and other exteroceptive sensations appear to have a bilateral representation, i.e., impulses from both sides of the body are received by each thalamus

Thalamic lesions

The effects resulting from a lesion involving the lateral thalamic nuclei will resemble those following an interruption of ascending pathways in the brain stem above their decussation. That is, all types of sensation on the opposite side of the body tend to be lost or grossly affected, or there may be severe pain as a result of destruction of inhibitory mechanisms or possibly of the direct stimulation (irritation) of pain fibers. A lesion involving the fibers after regrouping (i.e., after the finer discriminative types of sensation have been sorted from the cruder sensations and are ascending to the cortex) will cause a loss of the discriminative types of sensation alone. A lesion in the internal capsule (thalamic radiation) or corona radiata might have such an effect.

THE THALAMIC SYNDROME ("SYNDROME THALAMIQUE") of DÉJÉRINE AND ROUSSY. This is a characteristic picture occasionally seen in thalamic disease. The following are its chief features:

(1) Astereognosis and slight ataxia, due to the impairment of cutaneous and kinesthetic senses. These and the other effects given below are mainly on the contralateral side of the body.

(2) Some loss of tactile and thermal sensations over the body and face. The threshold for these sensations, and for pain, is frequently raised. Sometimes the patient with a thalamic lesion is unable with closed eyes to localize the position of a limb, and must grope in the air in order to find it (*thalamic phantom limb*), or he may have the illusion that the limb is not there at all (*amelognosia*).

(3) Spontaneous pain occurring in paroxysms and often excruciating. The pain may be so intense as to resist the action of powerful sedatives, e.g., morphine. A painful stimulus is felt much more acutely than is normal (hyperalgesia) and though, as mentioned above, the threshold for pain is often raised, the sensation when once aroused (by increasing the strength of stimulus) is excessively severe. The spontaneous pain and the magnification of the response to painful stimuli are usually referred to as the *thalamic over-reaction*.

(4) Various forms of paresthenia may be present.

(5) Hemiparesis and chronic or athetoid movements, but these are due to injury of neighboring parts.

As a result of their clinical researches, Head and Holmes suggested that the over-reaction phenomena are due to the release of the thalamus from the restraining influence which the cerebral cortex exerts normally through the corticothalamic fibers. According to this theory, the symptoms result from the interruption of the corticothalamic connections or their junction with cells in the posteroventral part of the lateral mass.

The thalamic syndrome includes other unpleasant sensations, even a form of stimulation, such as tickling the sole of the foot, which produces no discomfort on the sound side, is highly unpleasant on the other side. Many patients complain that shaving the affected side of the face feels as if the razor were being drawn over a raw surface. Even cutting the hair or nails may be objected to as being extremely disagreeable.

Sensations of pleasure may also increase in intensity. Warmth applied to the skin may cause the keenest enjoyment and evoke such expressions as "exquisite", "lovely", etc. The patient, however, may be quite unable to appreciate that the sensation is one of warmth. During emotion disagreeable sensations occur on the affected side. Thus one of Head's patients was so affected by music that he "could not stand the hymns on his affected side", another said that when the choir sang "a horrid feeling came on the affected side and the leg started to shake". In another patient pleasant feelings of a psychic nature were referred to the abnormal side. He said "I seem to crave for sympathy on my right side," and "My right hand seems to be more artistic."

THE HYPOTHALAMUS

The hypothalamus is the basal part of the diencephalon (interbrain). It lies in relation to the floor and lower parts of the walls of the 3rd ventricle (fig. 676). From a strictly anatomical point of view it may be taken to include the following structures: (a) *optic chiasma*, (b) *tuber cinereum* and the other nuclear masses in relation to the floor and ventral parts of the walls of the 3rd ventricle, (c) the *pituitary gland*, and (d) the *corpora mammillaria*. In physiological literature the term has usually a more restricted connotation including only (b) and (d) (see also p. 783).

Of the *nuclei of the hypothalamus* the greatest interest from the physiological point of view centers around the following

The *supraoptic nucleus* lies anteriorly, lateral to the optic chiasma and above the commencement of the corresponding optic tract. The *preoptic nucleus* is situated most anteriorly above and in front of the supraoptic nucleus, and immediately behind the lamina terminalis. The *paraventricular nucleus* is found above the supraoptic nucleus and is in close relationship medially to the wall of the 3rd ventricle. It is richly vascular, and its large vacuolated cells contain numerous granules, probably the mother substance of a secretory product. The *posterior hypothalamic group of nuclei* lie in the posterior part of the hypothalamus in relation to the wall of the 3rd ventricle, and include the mammillary nuclei. Cells of this group send fibers to the medulla oblongata and the lateral horns of the spinal cord, which constitute the spinal sympathetic center. The *tuber cinereum (nucleus tuber)* is a small eminence of gray matter situated at the base of the brain between the optic chiasma and the mammillary bodies, i.e., in the mid-region of the hypothalamus, from its vicinity grows the pituitary stalk. Two groups of cells can be clearly defined within it, they are known as the *dorsomedial* and *ventromedial hypothalamic nuclei*. The slight swelling caused by these nuclei and covered by the pars tuberalis of the adenohypophysis (p. 783), is called the "median eminence." The tuber cinereum is a center of the parasympathetic nervous system.

Fiber connections The hypothalamus receives fibers from the globus pallidus (pallidohypothalamic tract) and through the *medial forebrain bundle*, from the olfactory lobe and the parolfactory area. The medial forebrain bundle sweeps through the hypothalamus and in its course gives off fibers to several of the hypothalamic nuclei. The hypothalamus, especially the supraoptic and paraventricular nuclei, also receives fibers from prefrontal areas and the premotor area (6) of the cerebral cortex (p. 1037) both directly and indirectly through the thalamus. It also projects through the anterior and the dorsomedial nuclei of the thalamus to prefrontal areas. Other efferent tracts (sympathetic and parasympathetic) pass backwards in close relation to the ependyma of the third ventricle, and descend in the gray substance of the mid-brain (surrounding the cerebral aqueduct), pons, medulla (beneath the floor of the fourth ventricle) and spinal cord (sympathetic).

The various hypothalamic nuclei are in communication with one another through fiber tracts, the best known of these is the paraventricular-supraoptic tract, but others undoubtedly exist.

The supraoptic, and tuberal nuclei, and probably the paraventricular nuclei as well, are linked with the hypophysis by efferent fibers which descend the infundibular stalk (hypothalamo-hypophyseal tract).

The majority of these fibers terminate around the pituicytes of the pars nervosa but some can be traced into the pars intermedia.

Thus, there exists a pathway, through the hypothalamus, from the cerebral cortex to the secretory cells of the neural lobe of the hypophysis.

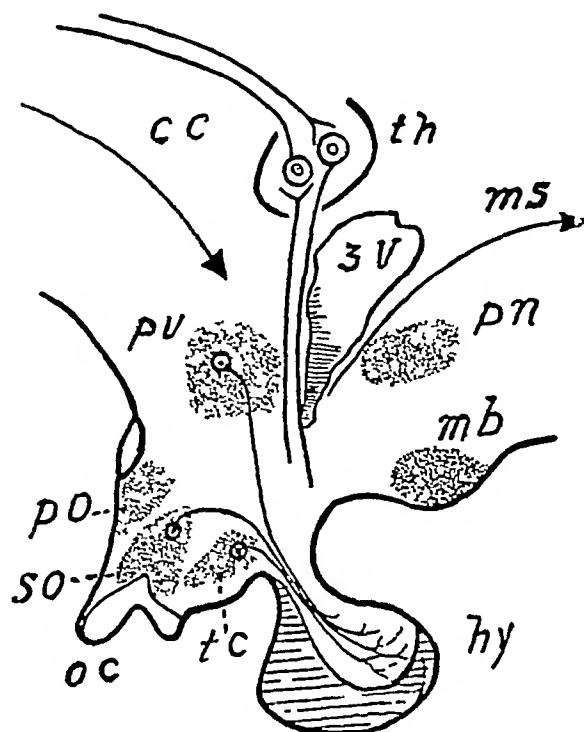


FIG. 67-6 Diagram of the nuclei and connections of the hypothalamus. *cc*, corticohypothalamic fiber, *pv*, paraventricular nucl., *po*, preoptic nucl., *so*, supraoptic nucl., *oc*, optic chiasma, *tc*, tuber cinereum, *hy*, hypophysis, *mb*, mammillary body, *pn*, posterior nucl., *ms*, fiber to midbrain and brain stem, *3V*, third ventricle, *th*, thalamus.

THE PHYSIOLOGY OF THE HYPOTHALAMIC NUCLEI

Modern experimental and clinical investigations have revealed the hypothalamus as a region of great physiological importance. Our knowledge of the functions of this part of the cerebrum is still, nevertheless, very incomplete. Much of the evidence is suggestive rather than conclusive and permits only tentative views to be held concerning many of its activities. It is the general opinion, however, that in this part of the diencephalon are contained the mechanisms for the control of certain primitive reactions (visceral and somatic) associated in animals with defense or attack, and in man with emotional states (fear, anger, etc.). This region is also believed to contain centers for the regulation of certain fundamental and vital processes, e.g., fat, carbohydrate

and water metabolism, and to exert a governing influence upon the body temperature, the gastric movements, the genital functions and the sleep rhythm (ch 69), hunger and thirst (ch 44)

It is now generally conceded that important centers governing the activities of the autonomic nervous system are situated in the hypothalamus. Those cell groups in the anterior part of the hypothalamus and in the tuberal region are believed to constitute a parasympathetic center, whereas from the posterior part a directing influence is exerted over sympathetic functions.

The pituitary and the nervous structures of the hypothalamus are intimately associated in function, indeed, they should be considered together as constituting a closely integrated neuro-glandular mechanism rather than as possessing distinct and independent functions. It will be recalled that the pars nervosa of the pituitary is developed as a downgrowth from the floor of the 3rd ventricle, and that the pituicytes are modified neuroglial cells. On the other hand, the cells of the supraoptic and paraventricular nuclei of the hypothalamus show evidence of secretory activity. They have not the appearance of typical nerve cells for they contain granules resembling those found in glandular cells and droplets or larger collections of a colloid material can be observed within their cytoplasm, they also show vacuole formation which is rarely encountered in nerve cells. The rich vascular network which surrounds the cells of these nuclei, capillaries in some instances actually penetrating their protoplasm, strongly suggests a glandular function.

A brief summary of the experimental evidence relating to hypothalamic functions will be given in the following paragraphs.

(1) Karplus and Kreidl were the first to furnish evidence of a sympathetic center in the hypothalamus. Upon electrical stimulation of this region they obtained pupillary dilatation, sweating and a rise in blood pressure. Inhibition of intestinal movements also results. Liberation of adrenaline has also been reported to follow stimulation of the hypothalamus. By means of needle electrodes inserted into the region of the lateral nucleus Bronk and his colleagues have recorded impulses from sympathetic efferent nerves during stimulation of the hypothalamus. On the other hand, rhythmical variations in potential were produced in the hypothalamus by the stimulation of certain afferent nerves through which reflex sympathetic responses may be elicited.

(2) Beattie, Brow and Long found that extrasystoles produced in the cat by means of chloroform anesthesia, and which had been shown by Levy to be dependent upon sympathetic impulses (p 237), were abolished by a destructive lesion placed in the posterior hypothalamic nuclei or by a section of the brain behind this region. Stimulation of the posterior hypothalamic region, on the other hand, caused extrasystoles to appear in an animal which previously had been free from these cardiac irregularities. Animals subjected to such lesions also showed hyperglycemia and glycosuria. Drowsiness for two or three days following the operation was a noticeable feature in some animals, they also showed a change in behavior, being more docile and "tamer" after the operation (see "sham rage" below). The nervous system of animals examined histologically some time after the lesions in the posterior hypothalamic nuclei had been made showed degenerating fibers which entered the mid brain and descended through the brain stem and cord. Those in the latter situation entered the lateral column of gray matter at different levels down to the 3rd or 4th lumbar segment.

(3) Decerebration by a section through the mid brain causes a profound fall in body temperature. No such loss of temperature control follows the removal of the cerebral cortex and thalamus, provided that the hypothalamus is left intact. It is significant that sympathetic effects, e.g., adrenaline liberation, ruffling of feathers or hairs, constriction of vessels and goose flesh, result from exposure to cold. Moreover, Cushing calls attention to the high temperature which frequently follows operations upon tumors in the region of the 3rd ventricle (see also p 737).

(4) Hess fixed electrodes in the hypothalamus of cats. After the animals had recovered from the operation, a weak electric current passed through the hypothalamus induced a state indistinguishable from normal sleep. He also states that ergotoxin injected directly into the hypothalamic region induces sleep. This observation is, however, difficult to reconcile with the fact mentioned below that the intravenous injection of ergotoxin produces sham rage. Sham rage and other manifestations characteristic of hypothalamic stimulation also follow the local injection of a minute amount of strychnine.

(5) Kabat and associates by means of an electrode fixed in the hypothalamus stimulated this region in the unanesthetized animal. Pupillary

dilatation, erection of hair, inhibition of gastrointestinal peristalsis, clawing and urination resulted. Stimulation of other parts of the brain produced none of these effects.

(6) The production of gastric lesions by hypothalamic injury and of ovulation by stimulation of the hypothalamus or of the hypothalamo-hypophyseal nerve tracts have been mentioned (pp 523 and 886). Evidence for the control of gastric motility by the hypothalamus has been cited on page 570.

Quasi-emotional state—"sham rage"

It was first demonstrated by Goltz that the reactions which usually accompany displeasure and anger are more readily evoked in an animal deprived of its cerebral cortex (decorticated) than in the normal animal. In Goltz's classical experiment the hemispheres and a large part of the thalamus were removed from a dog. The disposition of the animal was greatly altered by the operation, it being very readily aroused to anger. Barking, growling, baring the teeth or snapping occurred upon the least provocation. Cannon and Britton produced a similar state in cats by removal of the cortex, the decortication being performed by means of a pointed stilet inserted through the orbital cavities. Immediately following recovery from the anesthetic the animals showed the following remarkable phenomena which these observers termed "sham rage"—lashing of the tail, erection of the hairs, protrusion of the claws, dilatation of the pupils, sweating, struggling and biting, greatly increased rate of respiration and a rise in blood pressure. Most of these manifestations will be recognized as being of sympathetic origin. In many instances the picture is a combination of fear and anger. Liberation of adrenaline, as indicated by an increase in the rate of the denervated heart (p 835), also occurred. The mildest stimulus such as jarring the table or lightly touching the animal was sufficient to evoke a paroxysm of rage.³ In his decortication experiments Bard found that the posterior and ventral portion of the diencephalon was essentially responsible for the development of "sham rage". The typical quasi-emotional behavior occurred after decortication and section through the hypo-

thalamus at about the middle of the tuber cinereum. It also resulted from an operation which removed the corpora striata and the dorsal half of the diencephalon, i.e., the thalamus, but left the hypothalamus connected with the mid-brain. The condition failed to appear if the section separated the caudal part of the hypothalamus, i.e., the portion containing the posterior group of hypothalamic nuclei (p 1025), from the mid-brain.

Taking the results of decortication experiments as a whole, one is led to the conclusion that the activities of the hypothalamus are normally under inhibitory influences from the cerebral cortex and that "sham rage" is the result of the release of these primitive subcortical centers from higher control. Fulton and Ingraham, for instance, found that in order for the typical behavior to ensue decortication is unnecessary. Bilateral incision of the medial surface of the cerebral hemisphere in front of the optic chiasma is sufficient to produce hypothalamic over-reaction. Typical "sham rage" can also be produced in intact cats by the intravenous injection of ergotoxine. It has been suggested that certain conditions in man, e.g., the manifestations of fear in shell-shock, the unreasonable rage of drunkenness, and the emotional instability of certain mental derangements associated with degenerative changes in the cortex, may be due to the release of the hypothalamus from cortical control.

In more recent experiments, Bard and Mountcastle have shown that the restraining influence of higher levels of the cerebrum upon hypothalamic activity is not exerted by the neocortex, but by certain parts of the rhinencephalon (cortex of the pyriform lobe, amygdaloid nucleus and hippocampal formation) and an area of cortex on the medial aspect of the hemisphere lying ventral to the cingular sulcus in front and below the rostrum of the corpus callosum. This latter area of cortex is believed to be a transitional zone between the rhinencephalon and the neocortex. An operation which removed the neocortex alone, leaving the rhinencephalon and the area of transitional cortex uninjured, was performed upon cats whose behavior and temperament had been studied for some time previously. The manifestations of animals prepared in this way were the reverse of those which had been subjected to complete decortication. They showed extraordinary placidity. Responses expressive of pleasure were predominant and often exaggerated. Procedures, which in a normal animal provoke resentment, anger or even

³ Pseudoaffective states—displeasure, anger or rage, can be evoked in animals after a section caudal to the hypothalamus, but much less readily than after decortication, and they do not constitute the fully integrated response of the animal with hypothalamus intact.

rage, such as, pinching the tail strongly with surgical forceps, tying the animal down on its back, strong electric shocks applied to the skin, etc., evoked no sign of anger or even of resentment. The animals responded to mild nociceptive stimulation by purring or with other expressions of pleasure. An animal which before operation resented handling or petting, became affectionate, purred when stroked and submitted to being tied to the animal board "as though it enjoyed the maneuver." Removal of those parts of the rhinencephalon mentioned above or of the transitional area of cortex, transformed animals deprived of their neocortex, i.e., abnormally placid animals, into ones which, upon the slightest provocation, exhibited all the signs of rage.

It is apparent from these experiments that some part of the rhinencephalon and the transitional area of cortex exerts a restraining influence upon those hypothalamic mechanisms governing the expressions displeasure, resentment and anger. Since removal of the neocortex alone evokes expressions of a reverse nature, it would seem that the latter part of the brain tends normally to antagonize the restraining influence of the former part. This suggests that the temperament of an animal is normally dependent upon the balance struck between the activities of these two parts of the brain.

Emo or This is an appropriate place to consider the mental state with its accompanying reactions which is generally referred to as emotion. The word emotion is derived from the Latin meaning a "moving out." But there is an inward as well as an outward component of the emotional state. It can be analysed into subjective and objective elements—emotional feeling or experience, and certain visceral and somatic manifestations, e.g., pallor, blushing, cardiac acceleration, racial expression, etc.

According to the James-Lange theory, the emotional feeling is not aroused *primarily* in consciousness but is the result of the bodily reactions. Briefly, for example, we are frightened because the heart accelerates, the vessels constrict, the respiration quickens and the skeletal muscles increase their tone, or contract for purposes of defense, or in order that one may run away, the afferent impulses initiated by these various activities impinging upon consciousness arouse the feeling of fear. Sherrington showed, however, that the emotional state of a dog remained unaltered after a high spinal transection

and section of the vagus nerves, afferent impulses from the viscera and skeletal muscles being thus largely removed.

The modern view, which was advanced by Cannon and by Dana, proposes that emotional feeling, and the associated bodily reactions are the result of interaction between the cerebral cortex and the diencephalon—hypothalamus and thalamus (anterior nucleus)—visceral and somatic responses being *secondary* to the feeling of rage, delight, grief, etc., rather than the cause, which is dependent upon the cortex, probably on the orbital and cingular gyri and the hippocampus,⁴ but the emotional manifestations are initiated in subcortical levels. As we have seen, there is much experimental support for such a view. Also, clinically, outbursts of uncontrollable laughter or crying are sometimes associated with lesions of the diencephalon.

DISORDERS OF THE HYPOTHALAMUS

The effects which may result from lesions (e.g., tumors, encephalitis, etc.) involving the hypothalamic region fall into the following groups:

(a) disturbances in fat or carbohydrate or in water metabolism and (b) disorders of sleep, drowsiness, somnolence and, less commonly, abnormal wakefulness, or reversal of the sleep rhythm,⁵ (c) emotional manifestations, laughing, crying, or a state resembling "sham rage" in animals, may result, (d) phenomena attributable to sympathetic or parasympathetic stimulation, (e) disorders of the sexual functions.

Any one of the foregoing groups or effects may dominate the clinical condition to give rise to one or other of the following syndromes:

- (a) Diabetes insipidus
- (b) Dystrophia adiposo-genitalis
- (c) Laurence-Biedl-Moon syndrome.
- (d) Autonomic diencephalic epilepsy
- (e) Narcolepsy

These several conditions, with the exception of the last (e), have been considered in chapter 57.

Narcolepsy (see also p. 1068). This is the term applied to a disturbance in the sleep mechanism in which sudden attacks of an irresistible desire for sleep occur.

⁴ von Bonin speaks of a reverberating circuit—hypothalamus (mammillary body) to anterior thalamic nuclei via mammillo-thalamic tract, thence to anterior part of cingular gyrus, thence to cornu Ammonis, and finally back to the mammillary body through the torus and fimbria.

⁵ See Fulton and Bremer.

during the day-time. The duration of the attacks, which resemble normal sleep, is quite brief—from a few seconds to 20 minutes or so. It is only to such sudden and brief naps, and not to persistent drowsiness or to prolonged periods of pathological sleep that the term is applicable. Nocturnal sleep may be normal but it is often disturbed or there may be insomnia. Sleep may overcome the subject of narcolepsy while he is going about his usual occupation, while walking, in the middle of a conversation, during a meal, driving a car, etc. There may be many attacks during the day. A few cases were discovered during the war in soldiers under trial by court-martial for falling asleep on sentry duty. The condition may be a sequel to influenza or to epidemic encephalitis involving the hypothalamus or may result from a tumor or injury in this region. In other instances the condition appears without known cause (idiopathic narcolepsy). Though evidence is not conclusive, it is very likely that in these latter, also, disordered hypothalamic function is responsible, for other features, e.g., obesity, polyuria or impairment of the sexual functions, pointing to an abnormality of this region are frequently present. *Ephedrine sulphate* (25 mg three times daily) or *benzedrine* is used with benefit in idiopathic narcolepsy. *Cataplexy*⁶ is the term given to a condition allied to, and very frequently associated with narcolepsy, in which the patient as a result of some

emotion—amusement, anger, fear, embarrassment or surprise—is seized with complete helplessness. The attack is brief, lasting for a few seconds, or for a minute or two at the most. Consciousness is not lost but the muscles are completely toneless and powerless for the time, and if the attack supervenes while the subject is standing his knees fail him and he sinks to the ground. The deep reflexes are lost. A normal person may become “weak with laughter”, be “struck all of a heap”, or “transfixed” when surprised or shocked. Or his jaw may “drop” when confronted with some unexpected occurrence. Cataplexy is regarded as an exaggeration of this normal tendency, just as narcolepsy is an intensification of the desire of many normal persons to drop into a doze under certain circumstances. Mirth is especially likely to precipitate a cataplectic attack. One victim reported by Adie remarked, “At the scout’s camp the boys used to amuse themselves by making me laugh and then running away leaving me helpless on the ground.” Though narcolepsy occurs without cataplexy the converse is extremely rare. This association of the two conditions at once suggests a common pathogenesis, but the muscular atonicity characterizing the cataplectic attack cannot be explained upon any physiological basis. An interesting speculation has been made by Wilson, who compares the attacks to the defense reaction of certain animals whereby they fall into immobility when frightened, and suggests for them a certain biological significance, namely, that they are the relic of a primitive reaction uncovered by disease.

⁶ This should not be confused with catalepsy, an entirely different condition.

CHAPTER 68

THE CEREBRAL CORTEX THE PHYSIOLOGY OF SPEECH AND SOME OF ITS DISORDERS APRAXIA AND AGNOSIA EPILEPSY HEADACHE

MINUTE STRUCTURE OF THE CORTEX

The human cerebral cortex (or *pallium*) has a total area of about 220,000 square millimeters, not more than a third of this lies upon the free surface or crown of the convolutions. The remaining two-thirds of the gray mantle of the cerebrum occupies the walls of the sulci. The total number of nerve cells in the human cerebral cortex has been estimated at around 10^{10} . The number of fibers received from and projected to lower levels of the nervous system is in the neighborhood of 200 million. Added to these are fibers, many times more, which connect cells within the cortex (association fibers). On the basis of cellular structure the major part of the cortical gray matter is divisible into *six layers or laminae*. But these layers do not show identical histological appearances throughout the extent of the cortex. Characteristic differences in the depth of the individual layers and in their cellular components are found in the various regions. The six layers from the surface inwards with a general description of their cellular features as given by Economo¹ follows (see fig. 68 1)

I MOLECULAR (OR PLEXIFORM LAYER) In this, the most superficial layer, the terminal filaments of numerous dendrites from cells of deeper layers, as well as from the axons of Martinotti cells form a dense felted network. Its cells are sparse, they are small (4 to 6 μ) and pear shaped or fusiform.

II EXTERNAL GRANULAR LAYER consists of large numbers of small round, polygonal or triangular cells closely packed together. Their afferent processes pass into the overlying layer, their axons end mainly in deeper layers, but some enter the white substance of the hemispheres.

III PYRAMIDAL CELL LAYER Medium sized pyramidal cells are contained in the outer part of this layer, pyramidal cells of larger size and more sparsely distributed are present in the deeper part. It is customary, therefore, to subdivide this layer into an outer and an inner portion, Campbell refers to them as separate layers.

IV INTERNAL GRANULAR LAYER resembles the external granular layer in being composed of closely

packed masses of small stellate cells, but unlike the outer granular layer it is rich in nerve fibers. This layer contains many horizontal fibers which show as a whitestripe, or band known as the outer stripe or line of Baillarger, which is especially well marked in the calcarine cortex, but in this situation it is more usually referred to as the line of Vicq d'Azyr or of Gennari.

V GANGLIONIC LAYER (or internal pyramidal layer) consists of pyramidal cells of graded sizes. This layer is particularly well developed in the precentral (motor) cortex where giant pyramidal cells (Betz) are conspicuous. It contains, also, cells of Martinotti, these cells are peculiar in that their axons pass *outwards* toward the surface of the cortex and arborize in their own layer, or in overlying layers. Some of these cells can be found in nearly all layers of the cortex. Its deeper strata contain a dense network of fibers which forms the inner line of Baillarger.

VI FUSIFORM CELL LAYER, in contact with the white matter, is composed of closely packed small spindle shaped cells with their long diameters perpendicular. Cells of Martinotti and small stellate cells are present also in this layer but fewer in number.

It should be emphasized that the foregoing is no more than a general description of the histological structure of the cortex and that marked regional differences exist. Even the lamination itself is not a feature common to the entire cortex. In man one-twelfth of the cortical area shows little or no lamination, this portion, which is called the *allocortex* or *archipallium* comprises the cortex of the olfactory lobe (i.e., the pyriform area and the hippocampal, supracallosal and olfactory gyri, etc.) The laminated cortex, which in man constitutes the remaining eleven twelfths, but which in lower mammals is a much smaller fraction of the whole, is called the *isocortex* or *neopallium*.

The greater part of the human laminated cortex shows the cytoarchitectural features described above, and is therefore sometimes referred to as the *homotypical* cortex. In other more restricted areas, the cortex departs from the typical cytological appearance, chiefly in the preponderances of the small granular cells or of the pyramidal elements, and is called *heterotypical*. When pyramidal cells predominate, the cortex is called *agranular*, and when the small granule like cells

¹ For a detailed description, this author, or the earlier works of Campbell and of Bolton, should be consulted.

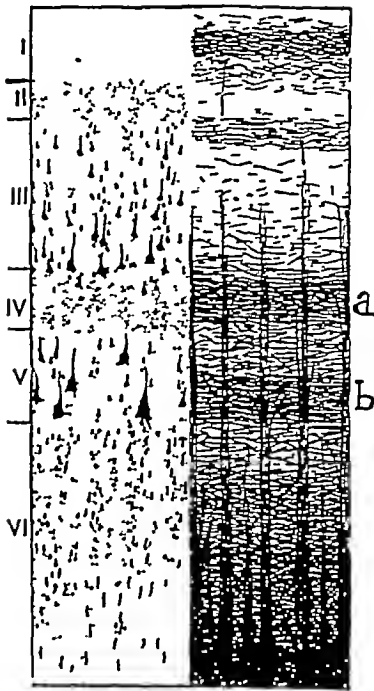
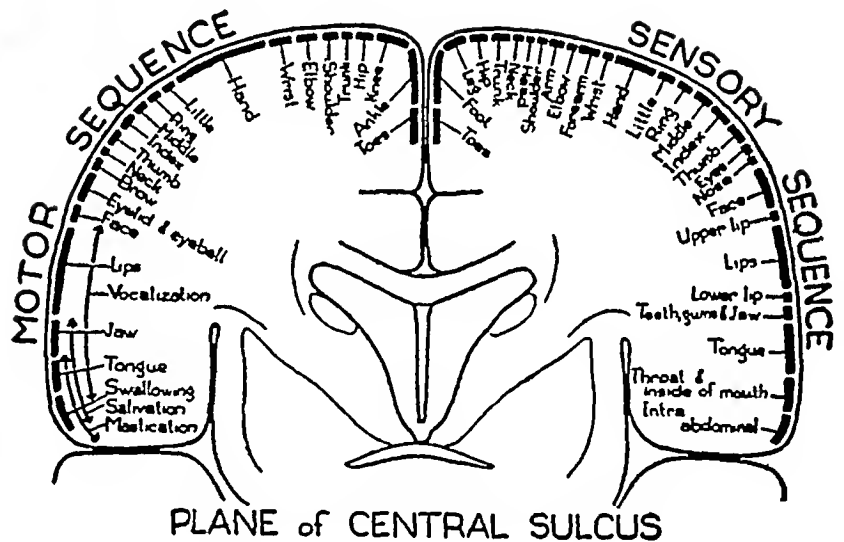


FIG 681 Showing layers of area 4 of human cerebral cortex (agranular), *left*, cell-bodies, Nissl stain, *right*, nerve fibers, Weigert stain, a, outer band of Baillarger, b, inner band of Baillarger. Note the large cells of Betz in the fifth layer. After Brodmann

FIG 682 Coronal section through the hemispheres to show the motor and sensory representations in the cerebral cortex (from Rasmussen and Penfield)



occupy most of the thickness of the cortex, with few pyramidal cells, it is called the *granular* cortex. The agranular cortex is thicker than the granular which is particularly thin at the frontal and occipital poles of the hemispheres.

The cortex of the precentral region (motor and premotor cortex) is of the agranular type. The granular layers (II and IV) are shallow, being encroached upon by the expansion of the pyramidal layers III and V. In the deep part of layer V of this area are situated the characteristic giant cells of Betz (60 to 80 microns in their longest diameters) (see below). These cells contribute fibers to the corticospinal (pyramidal) tracts (ch 66). The frontal cortex in front of the motor area

(i.e. in the premotor area), granular layers are also inconspicuous but the large cells of Betz are absent.

In the cortex of the walls of, and surrounding the calcarine fissure—the *area striata* or *visual area*—the outer and inner granular layers (II and IV) are expanded at the expense of the pyramidal layers (III and V). This very finely *granular* type of cortex, also called from its “dusty” appearance under the microscope the *komocortex*, is characteristic of sensory areas. It is present, though to a less fully specialized degree than in the visual cortex, in the post-central gyrus (somesthetic area) and in Heschl’s gyrus (auditory cortex, p 1044).

Upon gross examination of a section of the brain two lighter bands can be seen in the cortex

against the darker gray matter. These are produced by nerve fibers running parallel to the surface of the convolutions. They are known respectively as the outer (already mentioned) and inner bands of Baillarger. In the visual area the outer band is broad and prominent and is usually referred to in this part of the cortex as the band of Vicq d'Azyr or of Gennari, who had described it previously. The inner band is absent in this area.

LOCALIZATION OF FUNCTION IN THE CEREBRAL CORTEX

Campbell mapped the cortex into 20 areas which could be differentiated on the basis of their cyto-architectural characters. This number was raised to 47 by Brodmann, and subsequent workers (Vogts) have subdivided these again to make a total of 200.² Brodmann's areas will be given mainly here, but only those of interest from a functional point of view will be more than mentioned.

THE FRONTAL LOBE

The precentral cortex, areas 4, 4s, 6, 8 and 44

Area 4 is a tapering strip of agranular cortex lying with its wider end at the upper border of the hemisphere, in front of the central fissure (Rolandic), and occupying the posterior part of the precentral convolution (fig. 68.3). A good part of this area is buried, for it covers the anterior wall of the central fissure. It also turns over the upper border of the hemisphere and extends down the mesial aspect as far as the cingulate sulcus. Area 4, or as it is more generally designated, the *primary motor area*, is a center for voluntary movement. It gives origin to the pyramidal (corticospinal) tracts, and also projects to the pons (frontopontine tract), corpus striatum, red nucleus, and the subthalamus. Areas 4 and 6 are also connected by intracortical fibers. The muscles of the various parts of the body are represented in this area from the lowest point on the mesial surface to its lowest part in the lateral surface, in an order which is, in general, the inverse of that in the body itself, namely, toes, ankle, knee, hip, trunk, shoulder, arm, elbow, wrist, hand, fingers, brow, eyes, face, larynx, jaw and tongue (fig. 68.2). The parts of the face are not inverted. By suitable electrical stimulation of different parts of this area, movements of muscles on the opposite

² The microscopical study of the cortex and mapping it into areas according to the cytological characteristics is called *architectonics* or *cytoarchitectonics*.

side of the body can be evoked. Though the control is mainly contralateral there is also some ipsilateral representation. The evoked movements are discrete, and limited to small muscle groups, single muscles, or even part of a muscle. They are thus unlike the broader synergic movements elicited from the premotor area. *Muscles* rather than *movements* are represented in area 4. In man, nothing more complicated than a simple flexion was observed by Penfield upon stimulation of the motor area in conscious subjects. No acquired or skilled movement or any act which a newborn infant could not perform was ever evoked. The movements so induced are initiated in the 5th cortical layer, and presumably through the activation of the giant pyramidal cells of Betz, after destruction of the overlying layers by the method of thermocoagulation the motor responses cannot be elicited (Dusser de Barenne). It will be seen from figure 68.2 that the size of the cortical area of representation of a somatic part is proportionate to its functional activity and importance. For example, the area for the hand and fingers is many times greater than that for the much larger muscular bulk of the trunk. The prehensile tail of the spider monkey has a cortical area as large as that of the foot.

We have already seen that area 4 also contributes fibers to the extrapyramidal system, it has connections with the cerebellum and the posterior ventral nucleus of the thalamus, and is connected transcortically with area 6.

Vocalization has followed stimulation of points in the lower third or so of the precentral cortex, e.g. barking in the dog, in patients, a cry resembling that of an epileptic at the commencement of a fit, or of an infant can be called forth by electrical stimulation of this region (see Penfield and Bouldrey, Penfield and Rasmussen). But no sound even remotely resembling a spoken word has ever been evoked by electrical stimulation of this part of the cortex.

The supplementary motor area. This, in the human brain is a band of cortex on the mesial surface of the hemisphere lying in front of the primary motor area, and extending from the upper border to the cingulate sulcus. In man, stimulation within this area causes an isolated movement of a limb, a postural movement of the body (fig. 68.5, p. 1040), or rhythmical movements (Penfield and Jasper). Vocalization is also elicitable from the upper part of this area.

Area 4s This is a narrow band of cortex also known as the "strip" area of Hines, lying between the upper parts of areas 4 and 6. Its cytoarchitecture is of the agranular type, resembling that of area 4 except that the giant pyramidal cells are absent or very scarce. Electrical stimulation or strychninization of this area inhibits contractions elicited from area 4, and raises the threshold of excitation of the latter. Its ablation results in spasticity and an exaggeration of the myotatic

generalized in nature than those caused by the stimulation of area 4 and involve synergic groups of muscles. Though by suitable stimulation restricted movements can be elicited they are never as discrete and limited as those initiated from area 4. The movements elicited from the premotor area appear to have intent and purpose and are well integrated. They are slow in their onset, "build up" slowly, and outlast the stimulus (after discharge), which must be stronger than that

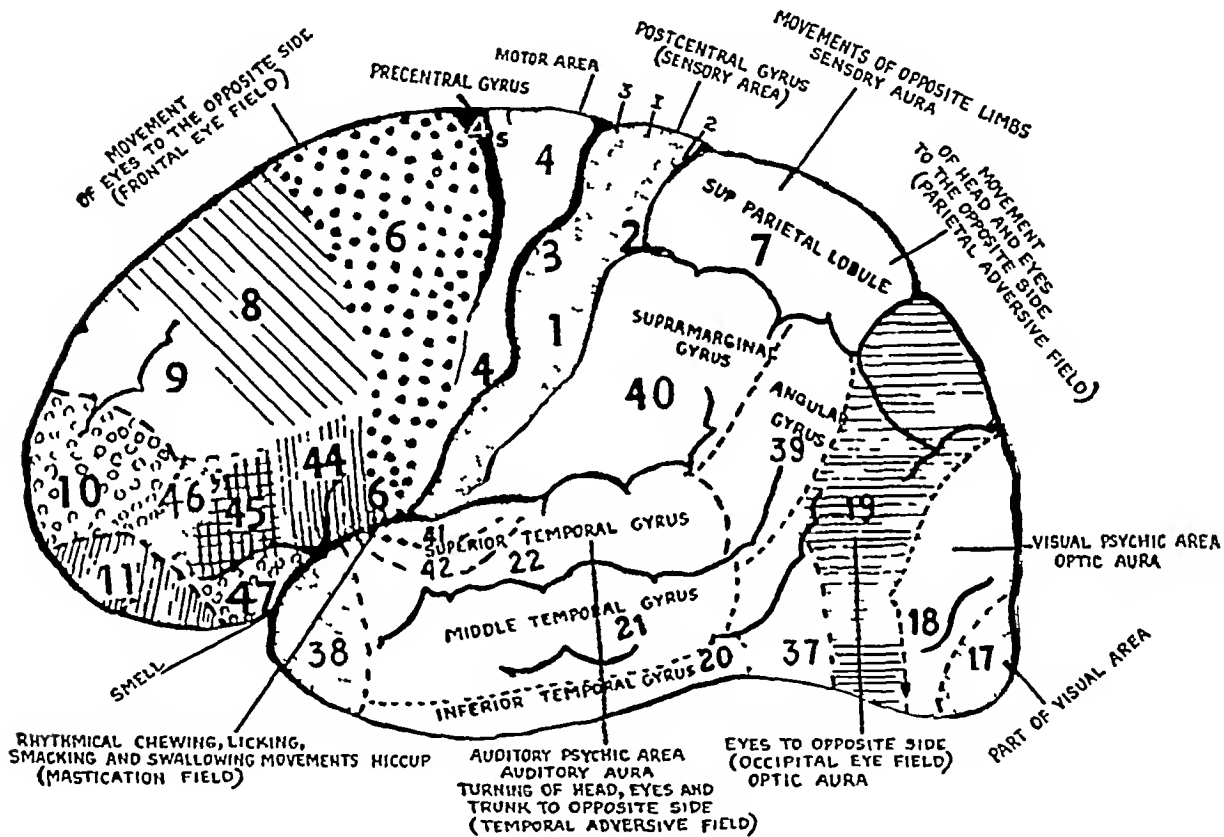


FIG 68.3 The cortex of the lateral surface of the hemisphere The numbers of areas are after Brodman

reflexes (see pp 960 and 1034) Area 4s projects to the corpus striatum (caudate nucleus) and through it to the thalamus, which connects again with area 4 It also sends fibers to the subthalamus and substantia nigra, and to the pontine nuclei, it contributes to the pyramidal tracts

Area 6 lies in front of areas 4s and 4. It is known as the *premotor area*. The topography of the premotor area corresponds in a general way to that of area 4. It is continued on to the mesial aspect of the hemisphere down to the cingular sulcus. It represents mainly the contralateral half of the body, but also to some extent the ipsilateral half. Area 6 differs functionally from area 4 in that the movements evoked by its stimulation are more

required for excitation of area 4. To this type of movement (which is of extrapyramidal origin) Hines has applied the term *holokinesis*³ as opposed to that initiated from the motor cortex (pyramidal type of movement) which she terms *idiokinesis*. A cut made between areas 4 and 6 abolishes the more complex movements elicitable from the latter, indicating that the two areas are connected

³ Stimulation of the precentral cortex of fetal monkeys removed by Caesarian section in the earlier days of gestation, before the pyramidal tracts have developed, gives only holokinetic responses. After birth as the pyramidal tracts become functional holokinesis gradually gives place to the idiokinetic type of movement, the former being evolvable only over the anterior part of the precentral cortex. After pyramidal tract section idiokinetic movement cannot be elicited.

by transcortical fibers, and that the impulses for the holokinetic type of movement are transmitted largely through area 4. After destruction of area 4 and the degeneration of its projections, or section of the pyramidal tracts, the usual result of stimulation of area 6 is inhibition, though some postural reactions may also occur.

Stimulation of upper part of area 6 causes adverse movements, turning of the head and eyes to the opposite side accompanied by extension or flexion of the contralateral limbs.

Vasomotor reactions whose topography corresponds approximately with the distribution of the motor responses can be evoked by stimulation of area 6.

The histology of the cortex of area 6 is similar to that of the motor area except for the important fact that the large pyramidal cells of Betz are absent from the fifth cortical layer. Smaller pyramidal cells, evidently motor in function, are seen in this layer. The efferent projections of area 6 are largely extrapyramidal, it is one of the principle sources of extrapyramidal fibers. It sends fibers to the corpus striatum, subthalamus, red nucleus, substantia nigra and pontine nuclei, they descend in the internal capsule. This area also receives projections from the posterior ventral nucleus of the thalamus, and from the cerebellum.

The effects of the ablation of area 4 and area 6

In lower vertebrates, e.g., reptiles, amphibians and birds, the rudimentary cortex plays a very subsidiary rôle in the control of motility. Subcortical and phylogenetically older portions of the cerebrum (corpus striatum and thalamus) are dominant (ch. 67). The frog after cortical extirpation behaves in an almost normal fashion, and even the dog or cat walks a short time after the cortex has been removed, most of the postural reactions being retained (thalamic animal). The motor area of these animals is much less easily excited by electrical stimulation than is that of the monkey, ape or man, and the excitable area is much smaller in extent.

In monkeys, as shown by Lashley, destruction of area 4 does not abolish the animal's ability to perform acts previously learned, such as opening a hasped box to obtain food.

In Fulton and Kennard's experiments on the chimpanzee, excision of area 4 (exclusive of 4s) was followed by a *flaccid* paralysis with *depression* of the tendon reflexes. The movements of the distal parts of the limbs, e.g., the fingers were most

seriously affected, manipulative dexterity was therefore greatly impaired, and never fully regained. The more proximal movements, e.g., the shoulder and elbow, recovered first, and eventually, almost completely. After this operation, the animal does not forget how to perform skilled acts which it has been taught, such as opening a box containing food, but is unable to execute the necessary manipulations or does so clumsily and with difficulty, owing to the paralysis of the finer movements. After destruction of area 4, a positive Babinski response without "fanning" of the toes was obtained. Rossolimo's sign was negative, vasomotor disturbances were absent. *In no instance was permanent spasticity, which has been usually considered in man to be due to a lesion of the motor area or corticospinal (pyramidal) pathway, observed after removal of the motor cortex alone.*

In contrast to the effects of ablation of area 4, excision of area 6 was followed by a plastic rigidity, increased tendon jerks, involuntary forced grasping and groping, and loss of the ability to perform learned acts because the memory of their patterns has been lost. Rossolimo's sign (p. 962) and fanning of the toes appeared, but not the Babinski response. Unilateral removal of the motor area (area 4) followed by excision of the corresponding premotor area resulted in *pronounced spasticity* and loss of voluntary power in the contralateral limbs, and a well-marked Babinski response *with* fanning of the toes. The condition resembled the hemiplegia which in the human subject results from a lesion in the internal capsule. Fulton and his associates conclude that the sign of Babinski without fanning of the outer toes (p. 1013) is indicative of a purely pyramidal lesion, the presence of the fanning component being an index of a premotor injury.

The main clinical features of the three types of lesion are shown in table 94.

The extrapyramidal connections of area 4 with the premotor area and with subcortical centers render it impossible by means of a cortical lesion to produce effects attributable purely to the destruction of corticospinal projections. But the corticospinal (pyramidal) fibers descend in two compact bundles—the pyramids—lying on the anterior aspect of the medulla (ch. 66). Marshall and later Tower sectioned the pyramid of one side in cats and Tower and Hines in the monkey, thus accomplishing the dissociation of pyramidal from extrapyramidal effects. Section of the pyramid of one side in the medulla at the level of the trapezoid

body resulted in paralysis of discrete, restricted movements of muscles, mainly on the opposite side, somewhat less in degree than that caused by ablation of area 4 itself. Larger movements (e.g., adhesive) and, with strong stimulation, epileptiform responses were still elicitable from area 6 as well as from areas 4s and 4. The most significant findings were the absence of spasticity of the affected muscles (indeed there was a certain degree of hypotonicity), a positive Babinsky without fanning of the toes, and some muscular wasting. Ablation of area 6, especially of its anterior part, after pyramidal section increased the paralysis and induced hypertonus of the muscles. Stimulation of area 4 caused inhibition of the extensor tone of

It can carry out independently of the motor area only relatively gross movements, e.g., adhesive and postural-primitive movements (holokinetic) of which a newborn infant is capable. (2) Areas 4, 4s and 6, exert an inhibitory influence through extrapyramidal paths upon movements (mainly contralateral) governed through the pyramidal tracts. Extrapyramidal projections

TABLE 94
Ablations

SIGNS	MOTOR AREA (4)	PREMOTOR AREA (6)	BOTH MOTOR AND PRE MOTOR AREAS
Babinski "extensor" response	+	0	+
Rossolimo	0	+	+
Fanning of toes	0	+	+
Spasticity	0	+	+
Vasomotor disturbances	0	+	+

the paralyzed limbs, as well as of the sound (i.e., ipsilateral) limbs. Stimulation of area 6, on the other hand, induced hypotonicity of the flexor muscles, the grasp reflex of the hand, for example, undergoing relaxation.

From the foregoing account of the various experimental results in this field the following inferences may be drawn: (1) Whereas the pyramidal system (area 4) governs the contractions of single muscles, parts of muscles, or small muscle groups, the extrapyramidal system, and especially projections from area 6, is concerned with the larger coordinated responses. The premotor area provides a background, or a broadly drawn pattern of muscular organization into which the pyramidal system fits the finer details. The premotor area is thus capable of synthesizing the small, discrete movements governed from the motor area into complicated purposeful acts, and is therefore essential for learning skilled manipula-

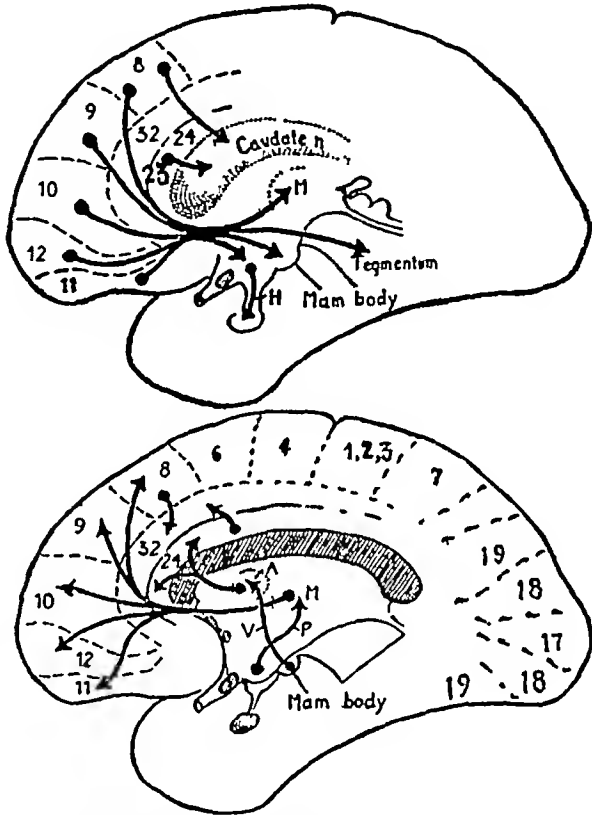


FIG 68-4 Connections of prefrontal areas (After Le Gros Clark, with minor additions). Lower sketch, right cerebral hemisphere from the medial aspect, showing afferent connections, A, anterior nucleus of the thalamus, M, dorsomedial nucleus of the thalamus, P, periventricular system of fibers ascending from the hypothalamus to the thalamus, v, mammillothalamic tract of Vicq D'Azyr. Upper sketch, efferent connections, M, dorsomedial nucleus of the thalamus, H, hypothalamohypophyseal tract.

from areas 4s and 6 inhibit the tone of the flexors (e.g., relaxation of the grasp of the fingers), those from area 4 inhibit the extensors. (3) Pyramidal function includes a tonic action upon the spinal centers. Interruption of pyramidal pathways results in hypotonicity of the paralyzed muscles, whereas lesions of extrapyramidal paths cause rigidity. These conclusions have important clinical implications. The spastic state of the affected muscles in hemiplegia, for example, has been generally considered to be indicative of a destruc-

tion of pyramidal pathways. It appears, however, that interruption of extrapyramidal fibers, in their course through the internal capsule in close association with the corticospinal tracts, is responsible for the hypertonicity of the muscles.

Fulton and his colleagues describe certain motor disorders in man which they consider to be characteristic of a lesion of the premotor area. These, which have been grouped under the term "*the syndrome of the premotor cortex*", are as follows: (a) awkwardness in the performance of skilled (learned) delicate acts, e.g., buttoning a collar, sewing, etc., but little early loss of power in the execution of gross movements, (b) spasticity and increased tendon jerks, (c) forced grasping and groping, late appearance of weakness of the hand-grasp and impairment of gross movements, and (d) autonomic disturbances. These effects are of course on the side of the body opposite to the lesion.

Grasping movements and tonic innervation

The weakness on the opposite side of the body which occurs in lesions of the premotor area is not uncommonly associated with a group of phenomena to which various terms have been applied, e.g., "forced grasping and groping", "grasp reflex", "tonic innervation" and so forth. A description of these phenomena follows.

(1) Merely touching the skin between the finger and thumb with a pencil results in slow flexion of the fingers. If the stimulating agent is withdrawn gently without disturbing the position of the patient's fingers no tightening of the grasp results but his hand and arm sometimes move through space (grobe) in the direction of the moving object, as if drawn by a magnet.

(2) *Grasp reflex*. When an object is placed in the paretic hand of the patient the fingers close slowly and gently around it, but any attempt made by the observer to withdraw the object often results in its being grasped more firmly. Nevertheless, when the patient clenches his empty fist he can relax the fingers again without difficulty.

(3) Any attempt to bend the arm or leg is met by an active resistance exerted by the antagonistic (stretched) muscles, this differs from the resistance offered by the ordinary spastic limb in that there is no sudden giving way with a "clasp-knife effect".

Walsh and Robertson have made a critical study of these phenomena and find that they are separable into two distinct components: (a) the *grasp movement* and (b) *tonic innervation*.

The *gentle* grasp and the groping movements just described (see (1) above) are voluntary and not reflex acts. That is to say, though they are automatic in nature and are taken to indicate deterioration of the psychomotor functions, the patient can prevent their occurrences if asked to do so. They disappear in stupor or in coma. The gentle grasping movement follows tactile stimuli alone or visual and tactile stimuli acting together. The groping movements can be elicited by visual stimuli alone but not by tactile stimuli alone.

Tonic innervation (see (3) above) is a stretch reflex (ch. 65). The *strong* grasp (2) which results when an attempt is made to remove an object from the hand is simply one phase—an incident—of this reflex. It is quite distinct from the gentle grasp movement. It results from the passive stretching of the flexor muscles of the fingers caused by the observer's attempt to extricate the object. The patient cannot relax the grasp and release the grasped object. It occurs in the unconscious patient. The grasp reflex is most pronounced when the patient is on his side and the affected arm uppermost (Fulton).

Richter and Hines produced the "tonic grasp reflex" in adult monkeys by removal of the premotor area. Excision of the motor areas or the prefrontal areas from both hemispheres did not cause the effect. A similar reflex is present in normal infants, Robinson showing that in these it is sufficiently strong to suspend the body from a bar for 2 minutes. A grasping reflex has also been described for the foot in lesions of the premotor cortex, it occurs in the normal infant up to the end of the first year and is said to occur in 50 per cent of Mongolian idiots. It is elicited by stroking the sole.

Area 8 (subdivided by the Vogts into sections α , β and γ) is called the *frontal eye field*. It lies in front of area 6. Stimulation of the cortex here causes a conjugate movement of the eyes to the opposite side, opening and closing of the eyelids, and sometimes dilatation of the pupils and lachrymation. No visual hallucinations occur in man, but an epileptiform seizure may be provoked which spreads to the adversive field (area 6).⁴ Ablation of area 8 of one side in monkeys causes the eyes to be turned up toward the side of the lesion, and temporary paralysis of the conjugate eye movement. The animal in walking circles

⁴ Stimulation of area 8 in man also causes conjugate deviation of the eyes and turning of the head to the opposite side.

toward the side of the lesion. No abnormalities are observed in the pupils or in the movements of the lids. Area 8 has depressor function. This area receives fibers from the dorsomedial nucleus of the thalamus, through which it is connected with the hypothalamus. It also connects by long association tracts, both afferent and efferent, with the occipital lobe (visual area, 18) and with the tegmentum of the midbrain (probably with the oculomotor nuclei).⁵

Area 44 in the region of the posterior part of the frontal operculum in the *dominant hemisphere* (which is the left in right-handed persons) is the motor area for speech (Broca's area). Stimulation of this area in conscious patients causes their speech to be abruptly arrested.

PREFRONTAL OR ORBITOFRONTAL REGION

The prefrontal region embraces the part of the frontal lobe in front of areas 8 and 44, but it includes the orbital as well as the lateral aspect of the lobe, and for this reason is also known as the orbitofrontal region. It is divided into the areas, 9, 10, 11, 12, 13,⁶ and 14. The orbital surface is occupied by parts of 10, 11 and 12, and all of 13 and 14. These areas until recent years have been thought to be inexcitable, and were therefore referred to as "silent" or association areas. But upon stimulation by a suitable electric current having pulses of low frequency, autonomic, respiratory, circulatory, renal (ch 35), and gastrointestinal responses can be elicited. After bilateral removal of these areas the blood pressure is reduced, gastric secretion suppressed, and gastrointestinal movements increased. In man, suffering from intractable pain is relieved. These areas receive important fibers from the dorsomedial nucleus of the thalamus, which in turn receives impulses from the hypothalamus. These orbitofrontal areas, especially those of the orbital surface, are thought also to be closely associated

⁵ Unilateral ablation of this area in monkeys is followed by a visual defect in the form of failure to recognize objects in the opposite homonymous halves of the visual fields—a *pseudohemianopia*—results. When area 8 is destroyed on both sides the animal does not react in a normal manner to visual stimuli. It may appear to be blind for it walks into or stumbles over obstructions in its path, and tends to stare straight ahead with an immobile "wooden" expression (see Kennard and Ectors). Yet, an animal will follow an object with its eyes and will seize anything offered to it though failing, apparently, to recognize it or to understand what to do with it.

⁶ Area 13, so designated by Walker, corresponds to Brodmann's 47.

with emotional feeling. Their part in the scheme of cortical representation of sensation is uncertain, but they are probably concerned with visceral rather than with somatic sensations through the hypothalamus and dorsomedial thalamic nucleus.

Area 24, anterior part of the cingular gyrus on the mesial aspect of the hemisphere, is a powerful suppressor area. Autonomic responses following stimulation of this area are pupillary dilatation, pilo-erection, acceleration of the heart, and a fall in blood pressure. Arrest of respiration in expiration also results.

The prefrontal areas have extensive subcortical and cortical connections. Areas from 9 to 12 in the monkey receive fibers from the dorsomedial nucleus of the thalamus. Fibers ascending to the orbital areas 11 and 12 to this nucleus have also been demonstrated in man from a study of retrograde degeneration in the nucleus as a result of leukotomy (p 1038). The anterior nucleus of the thalamus also sends fibers to the prefrontal area (area 23), and both dorsomedial and anterior thalamic nuclei receive fibers from the hypothalamus. The pathway from hypothalamus to the cortex through the dorsomedial nucleus was established by stimulation of the posterior hypothalamic region and observing the action potentials in the thalamic nucleus and the cortex. The greater part of the prefrontal cortex is considered by Le Gros Clark as a projection area for the hypothalamus, just as the occipital cortex and auditory cortex are projection areas, respectively, for the retina and the cochlea (Fig 68 4).

The prefrontal areas are connected by efferent fibers with (a) the hypothalamus through the dorsomedial nucleus of the thalamus, (b) there is also good evidence for the existence of a direct efferent connection with the hypothalamus both in animals and in man, (c) with the caudate nucleus, (d) nuclei of the pons (from area 10), (e) tegmentum of the mid-brain.

Unilateral or bilateral removal of the *prefrontal area* does not cause paralysis either in the monkey or in man. Unilateral removal of the human prefrontal area is without any outstanding effect. The mental processes are impaired only to a minor extent. Some loss of initiative and mental alertness, and lowered ability for arithmetical calculations may be the sole result of the operation. Memory, judgment and intellect often show little or no deterioration. Removal of the prefrontal area of the dominant hemisphere (i.e., the left in right-handed and the right in left-handed persons) tends to produce somewhat greater alterations in character or intellect than does a similar operation upon the non-dominant side. But even the bilat-

eral excision of prefrontal areas is followed by a surprisingly moderate mental defect. In an operation for the eradication of a tumor Dandy excised the frontal lobes on both sides in front of the premotor areas (reported by Brickner). The subject of this extensive extirpation appeared of normal intelligence upon a casual acquaintance. It is reported that for an hour he toured the hospital with two visiting neurologists who failed to notice in him any mental abnormality. A more intimate knowledge of the patient, however, revealed very definite defects of character and mentality. His mental age was about thirteen years, his intelligence quotient eighty. The main features shown by this subject and which may be taken generally as representative of the effects of extensive prefrontal destruction are listed below. They constitute what has been called the *frontal lobe syndrome*.

(1) *Lack of restraint* leading to boasting, self-aggrandizement, hostility, aggressiveness

(2) *Distractibility and restlessness*—difficulty in fixing attention

(3) *Hypermotility* which appears to be due to the loss especially of area 13

(4) *Flight of ideas*, puerile fantasies, emotional instability, facetiousness, punning

(5) *Lack of initiative*, and difficulty in planning any course of action

(6) *Impairment of memory*, for recent events but not for remote events

(7) *Impairment of moral and social sense*, loss of love for family

(8) *Failure to realize*, or indifference to, the seriousness of his condition, and a sense of well-being (euphoria)

Among some of the other manifestations which may follow a prefrontal defect are (a) increased

The classical example of a severe prefrontal injury which resulted in surprisingly little mental defect, is the case of the American, Phineas Gage (1848), who suffered extensive damage of his frontal lobes by an iron tamping bar driven through his head by an explosion. The bar penetrated the left orbit and emerged from the midline of the head just anterior to the coronal suture. He was stunned for only an hour and was able with assistance to walk to a surgeon's office. He lived for 12 years after the accident and showed in general these mental changes. A case has also been reported in which the prefrontal cortex on both sides were absent or degenerated. The condition existed from early childhood yet, at adult age the subject of this defect, after a careful psychologic and psychiatric examination, was reported to be of normal intelligence, though some defect in planning capacity and in the ability "to organize his behavior toward a relatively remote goal" were revealed.

appetite, in spite of which (and as a result of the *hypermotility*) there is loss of weight, (b) *impaired control of the sphincter of the bladder or rectum*, (c) *disturbances of orientation in time and space*, and (d) *tremor*

Chimpanzees which have had both prefrontal areas removed show restlessness and are easily distracted, though they remain alert and evince a keen interest in things around them. Fulton and Jacobson reported in 1935 that after this operation chimpanzees failed to show temper tantrums and other neurotic effects of frustration (experimental neurosis, p 1062). This observation formed the basis for the surgical treatment of certain psychoneuroses in man, first employed by the Portuguese surgeon, Egaz Moniz in 1936. The operation, called frontal lobotomy consists in severing the fibers connecting prefrontal areas with subcortical centers (probably the thalamus and hypothalamus). Excision of the orbital surfaces of the frontal lobes (bilateral orbital gyrectomy) which has also been undertaken for anxiety and depressive states, relieves intractable pain.

The frontal lobes and intelligence The development of the frontal lobes bears in general a direct relationship to the level of an animal in the phylogenetic scale and to its intelligence. This has led to the belief that this part of the cerebrum is the seat of the intelligence of animals and the "center" or "organ" of the mind of man. Within this region those processes underlying intellectual, as well as moral and emotional attributes were supposedly carried out. The absence of any overt effect from stimulation of these so-called silent areas seemed to confirm this belief. Extirpation experiments and lobectomy or injury in man, however, show decisively that the prefrontal area cannot be looked upon as a region where these higher mental qualities reside exclusively or even predominantly. Intelligence depends upon a knowledge of the external world received through various channels. Visual, auditory, somesthetic perceptions, etc., are received and stored as memories in cortical areas situated in the occipital, temporal and parietal lobes. Tracts of association fibers in turn link together these several primary areas, sensations of various types are thereby brought into relationship, and synthesized into more complex memories. Thus, as time passes, the fabric of our experience is woven in patterns of greater and greater intricacy.

The progressive increase in size of this part of

the brain through the upper levels of the phylogenetic scale is not as great as that of the parietal lobe, which supports the conclusion derived from extirpation experiments that the prefrontal areas are not especially related to intelligence, nor to memory, upon which mental ability so largely depends. The lower posterior part of the parietal lobe is more closely associated with these functions of the brain, defects of memory commonly follow injury to this part of the parietal lobe or to the temporal lobe.

It is probable that the prefrontal area merely represents a region of relatively high associative or synthetic capabilities. The subcortical connections of this area, as well as the effects upon behavior resulting from its injury, strongly suggest that it is concerned with emotional feeling. After its bilateral removal the cerebrum deprived of the synthesizing faculty of this region is incapable of the more elaborate association of those experiences required for the formulation of abstract ideas and more accurate judgment, and for the guidance of conduct in conformity with social customs. Here also are mainly located those mental processes relating to "prediction", forecasting, or to any planned action. A person deprived of these areas would have little ability as a strategist, even a housewife after the loss of these parts of the cortex would experience some difficulty in planning a meal. Nevertheless, synthesis at somewhat lower levels is still possible of achievement through the remaining cerebral tissue. Mental capacity according to this conception is therefore a function of the cerebral cortex as a whole rather than of any particular region. The *quantity* of tissue removed rather than its *location* is the more important factor in determining the degree of mental impairment which will result. (See Lashley.) It may be mentioned in this regard that the symptoms which are observed in prefrontal lesions can all be accounted for by a reduction in associative ability. Bolton has also shown that in amentia and dementia the degeneration of the cortical layers is not localized to any particular region but is distributed over the hemispheres.

The seat of consciousness The cerebral cortex has long been assumed to be the "organ" or locus of consciousness, a general preference being held for the prefrontal areas. But Penfield and Jasper offer a fresh conception of the neural basis of consciousness—that it is seated in the thalamus, hypothalamus and neighboring parts of the upper

brain stem. They speak of these structures and their connections as the centrencephalic system, and consider the latter rather than the cerebral cortex as "the highest level of neuronal integration." Some or other manipulation in the region of the medial part of the thalamus during an operation, as in an instance reported by Penfield, may be followed rather suddenly by loss of consciousness which is soon regained when the procedure which precipitated the lapse is stopped. According to this conception, it is the function of the neopallium to elaborate upon the field of consciousness for which the centrencephalic system provides the primary integrative basis. The cortex, of course, is essential for the development of those attributes peculiar to the brain of man—abstract thought, speech, mathematical calculation, imagination, creative ability, and all those higher mental capacities which come within the realm of intellect, and which are so variable from person to person.⁸

Facilitatory and suppressor bands and tracts When weak electrical stimuli are applied at one second intervals to area 4, it is found that the first stimulus of the series causes a lowering of the threshold for those which follow, or a weak stimulus which alone is ineffective may, if one of a series, become effective, neighboring neurons, as a result of facilitation, being brought into action (recruitment). After their excitation the threshold of these neurons is raised, their activity is suppressed. These are examples of *local* facilitation and suppression, i.e., the effects are produced within the area in which the stimulus is applied. Facilitation and suppression may also be induced upon one area by stimulation of another, either electrically or by the application of strychnine. Thus stimulation of the premotor area (area 6) or postcentral gyrus (1, 3, 2) causes facilitation in area 4, while suppression of activity in area 4 and other areas is produced by stimu-

⁸ The following is a quotation from Penfield and Jasper's recent monograph, "It is obvious that the higher mental functions which distinguish man from lower animals, such as speech, the capacity for higher mathematics, and other abstract thought processes, are not possible without the cortex, particularly that in the temporal and frontal lobes. The vast interconnected network of cells and fibers in the cortical matrix must, therefore, constitute an essential part of the machinery of the mind. But without the constant selective activating influences of the reticular network of the higher brain stem, the cortical mantle lies dormant. Without an integrated system for control of excitatory and inhibitory effects upon local cortical functions of the two hemispheres, co-ordination of cortical function as a whole would be impossible. Highest level functions cannot be strictly localized, but result from a dynamic interaction between centrencephalic mechanisms and those areas of cortex the function of which is momentarily being employed at a given time."

lating 4s. These are examples of *disinhibition* facilitation and suppression. The suppressor action from 4s depressor strip causes (a) inhibition of existing contractions on the opposite side of the body, (b) a rise in the threshold of area 4 and (c) cessation of its spontaneous electrical potentials. The suppressor action is not mediated transcortically but through a subcortical circuit—to caudate nucleus, to globus pallidus, to ventrolateral nucleus of thalamus (which is inhibited), and thence to area 4 via thalamocortical connections. The spontaneous electrical potentials ordinarily occurring in the striate structures are weak, but after cortico-subcortical connections are severed, i.e., the circuit is interrupted, bursts of high-voltage waves appear.

Several other cortical depressor areas have since been discovered besides those mentioned, especially in areas 7, 8, 12, 19 and 24 (7s, 8s, 12s, 19s, and 24s).

Descending tracts from the cortical depressor bands receive an accession of fibers from the caudate nucleus, and continue to a depressor center in the reticular formation of the medulla. From here via the reticulospinal tract an inhibitory influence is exerted upon stretch reflexes and decerebrate rigidity as well as upon movements induced by cortical stimulation. This depressor area in the medulla also receives impulses from the cerebellum. In a more lateral position in the reticular formation, and extending upwards into the midbrain lies a facilitation tract which enhances the activity of the spinal motoneurons acting in conjunction with the vestibulospinal tracts in maintaining the stretch reflexes which are the basis of decerebrate rigidity. The latter is less pronounced after destruction of this tract has been interrupted whereas it is increased after destruction of the medullary depressor area. The cortical depressor areas are gradually inactivated by cyanide, decerebrate rigidity then gradually appears as a result of the unbalanced action of the facilitation tract. These investigations have provided an explanation for the fact that large lesions in the pons may produce muscular hypotonus rather than rigidity as a result of the interruption of the facilitation tract.

THE PARIETAL LOBE

The somesthetic (somesthetic) area

The *postcentral gyrus*, i.e., the band of cortex lying behind and including the posterior lip and

* Attempts by Penfield and associates to demonstrate depressor areas in the cortex of patients have failed repeatedly, and it is doubtful that they exist in the human brain. Some have questioned even that *specific* cortical areas, i.e., areas whose function is confined to inhibition, exist in animals, and do not believe that the effects which have been observed represent a physiological function. Jasper, for example, has been unable on a number of occasions to inhibit movements by stimulation of the so-called suppressor areas on the lateral aspect of the cortex of cats or monkeys, and Clark and Ward obtained, as a rule, tonic movements from the stimulation of area 24.

wall of the fissure of Rolando (areas 3, 1 and 2) is sensory in function and is known as the *somesthetic area*. This band of sensory cortex turns over the upper border of the hemisphere, and extends down the mesial surface as far as the cingulate gyrus. As in the case of the primary motor area, the sensory areas of the body are, in a general way, represented mainly in inverted order, from the lowest part of the mesial surface to a corresponding part of the lateral surface. Thus, the area for the toes is at the lower part of the former surface, that for the leg near the upper border of the hemisphere, while the face, mouth and tongue areas are found in the lower part of the lateral surface. The parts of the face, however, are represented in uninverted sequence—brow, eyelids, nose, lips, in this order from above downward (fig 68.2). Though broadly

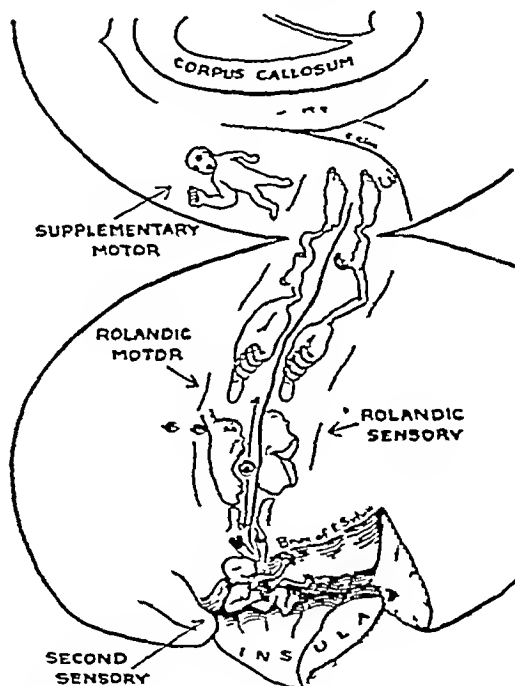
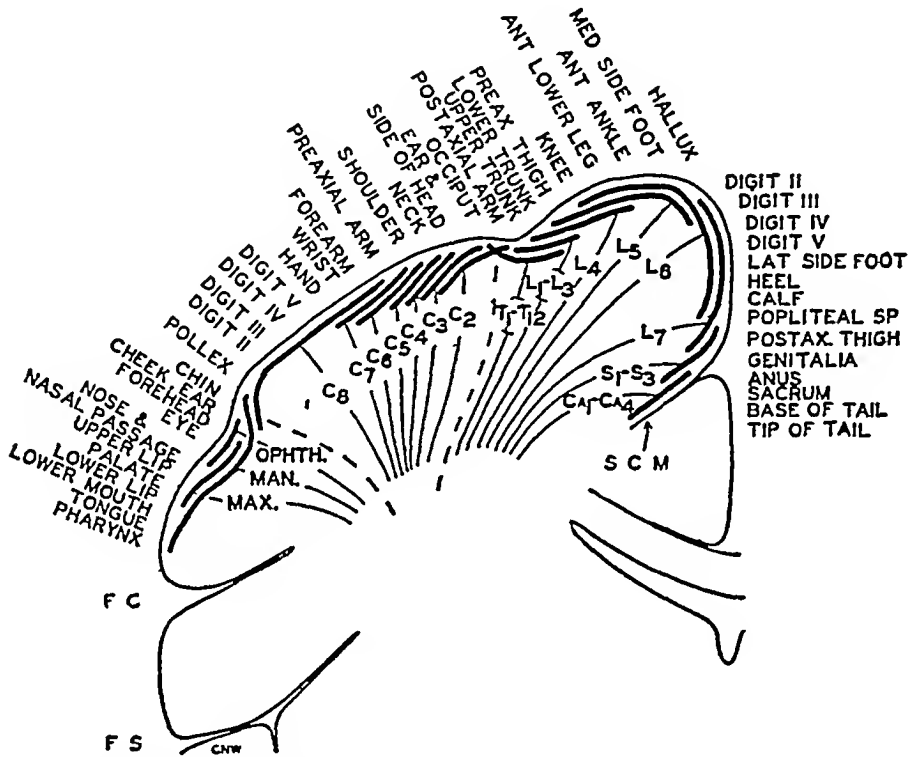


FIG 68.5 Showing figurines of motor and sensory representation in the cerebral cortex, slightly modified from Penfield and Jasper, *Epilepsy and the Functional Anatomy of the Human Brain*, courtesy of Dr. Penfield.

speaking, sensory representation in the cortex has this regional representation, when the topography of the different parts is examined more in detail it is found that the order of cortical representation conforms closely to that of the dermatomes, i.e., to the spinal innervation. For example, the cortical sequence for the upper limb (monkey) is postaxial arm, occiput, ear, side of head, neck,

The sensory area of the cortex is not confined to the postcentral gyrus but extends forward into the precentral gyrus. Penfield and his associates found that 25 per cent of stimulations of the latter area in patients caused a sensation, with or



Marshall, Woolsey and Bard have used a method to map out the cortical representation of tactile sensibility in monkeys, based upon the fact that impulses set up by stimuli applied peripherally can be recorded as action currents

from the surface of the hemisphere (fig 687) This is known as the reversed or crossed projection. They found tactile sensations represented contralaterally in the postcentral gyrus (areas 3, 1 and 2) Ipsilateral representation was not observed, except for a part of the face (e.g., lips, tongue and lower cheek). In no instance was evidence obtained of the precentral representation of this sensation.

As in the motor area the extent of the sensory representation in the post-central gyrus is greatest for those parts of the body which are of most importance in acquiring information concerning surrounding objects. Thus, in man and in the monkey, the cortical area for the hand and arm is larger than that for the trunk or leg, while in such animals as the rabbit, cat or pig the area for the face, lips, snout and vibrissae is much larger

tory areas are employed, namely auditory and visual areas I and II. Somatic area II lies in the upper wall of the Sylvian fissure, that is, below the face area of somatic area I, which is much larger in extent. The order of representation in sensory area II is the inverse of that of sensory area I, the face being above the arm in the upper part of the wall of the fissure and the representation of the foot at the bottom. The significance of the dual cortical representation is unknown. Area II is thought to be the more primitive of the two, that is, to have been developed at a much earlier phylogenetic period. It is possible that it receives fibers subserving the cruder forms of sensation (the protopathic system of Head) while the primary area receives impulses upon which more crucial perception (epicritic) is based (fig 685).

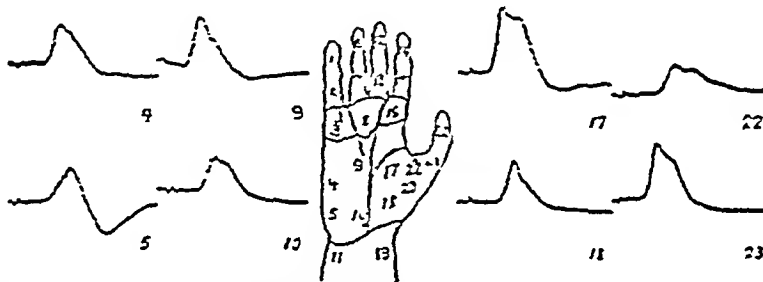


FIG 687 Action potentials recorded from postcentral gyrus during tactile stimulation of points on palm. (Part of figure from Marshall Woolsey and Bard.)

than that for the parts. In the pig the representation for the snout takes up a large part of the post-central gyrus, whereas that for the feet is insignificant. In the Shetland pony the cortical area receiving impulses from the nostril is nearly as large as that representing the rest of the body surface. The large facial representation in man appears to be an exception for, after infancy, we do not gain information by feeling objects with the lips or by putting them into the mouth. It is looked upon, however, as an inheritance from our animal ancestor.

Somatic sensation as well as the visual and auditory senses have a double representation in the cortex of each hemisphere. This dual somatic representation was first observed by Adrian, in the case of the cats of the cat. The somesthetic area described above and long recognized, is now called the "primary" sensory area, or, as Woolsey has suggested, somatosensory area I. The more recently discovered sensory area is called somatosensory area II. Corresponding terms for the dual visual and audi-

All somatic sensory impulses ascend in the medial, spinal and trigeminal lemnisci to the thalamus, which is the subcortical destination of crude sensations. Though rather vague sensations described as tingling numbness, tickling, pricking, or as the movement of a limb, are aroused by stimulation of the postcentral gyrus and are referred to some part of the body depending upon the area stimulated, pain is never aroused and rarely even discomfort. There does not appear to be an area of the cortex essential for the perception of pain for "no removal of the cortex anywhere prevents pain from being felt" (Penfield). Impulses for pain reaching the thalamus can arouse this feeling without being transmitted to the cortex. Though, as mentioned earlier, severe intractable pain usually associated with disease of internal structures is relieved by the excision of orbital areas, the perception of pain is not abolished, but the patient ceases to suffer and complain, he no longer worries about it.

The finer sensations of touch and temperature

and the sense of position and movement are relayed to the cortex by fibers which ascend through the internal capsule and corona radiata. The function of the somesthetic cortex, according to the theory of Head and his associates, is not, however, simply to record these several primary sensations. Its activity lies in the psychic sphere, cortical sensation, to quote these observers, is one of the "elementary processes of the mind." The somesthetic area brings its discriminative and synthesizing abilities to bear upon the primary sensations which it receives, and from these are formed our perceptions of the qualities of external objects, such as their size, shape, weight, texture, etc., and of the positions of our limbs in space. Through cortical activity a particular sensation is subjected to critical appraisal and compared with or related to another simultaneous or consecutive sensation, thereby, its intensity and nature are accurately judged. Thus, according to Head, the somesthetic area of the cortex, through the integration of the primary sensations, becomes endowed with three discriminative faculties.¹⁰ These are

(a) SPATIAL RECOGNITION—the appreciation of relationships in space, e.g., the recognition of position and passive movement of the limb, the discrimination of two points, and the localization of a point which has been touched

(b) RECOGNITION OF THE RELATIVE INTENSITY of different stimuli, e.g., that one object is warmer or cooler or more intense than another

(c) RECOGNITION OF SIMILARITY AND DIFFERENCES—appreciation of the shape, relative size, and texture of objects, and the estimation of their weights—*stereognosis*

In a lesion involving the somesthetic area, one or other, but usually all three, of these faculties are disturbed. Spatial recognition shows the greatest disturbance the farther forward the lesion lies in the somesthetic area. Appreciation of intensity is disturbed most by lesions involving the foot of the postcentral gyrus and the supra-marginal and angular gyri. Recognition of similarity and difference is affected most by lesions of the middle of the postcentral gyrus.

In the lesion confined to the cortex, the fibers ascending from the thalamus and conveying the

¹⁰ According to the conception of Penfield and Jasper, impulses received in the sensorimotor cortex return to the thalamus and other parts of the centrencephalic system where the highest level of functional integration is situated (p. 1039)

primary sensations, e.g., light touch, temperature, passive movement and position, etc., upon which the cortical faculties depend, are intact and the sensations are appreciated. The subject of such a lesion has difficulty, however, in bringing the necessary discriminative ability to bear upon the sensation in order to judge it, and he is unable to synthesize different sensations into a composite impression which will enable him readily to identify an object. When tested, he is uncertain in his answers, which tend to vary from moment to moment, and it is difficult for the examiner to determine the threshold for a given sensation. For example, though he recognizes that an object is warm, he cannot say whether it is warmer or less warm than another object which he has felt previously or at the same time. He responds to tactile stimuli, but also with inconstancy, and he is often even less consistent in his answers when the strength of the stimulus is increased. He cannot locate the point touched and may respond when not touched (hallucination of touch). The weights of objects placed upon the hand cannot be estimated, and a fabric (e.g., silk or tweed), though felt to be smooth or rough cannot be recognized.

Hypotonia may also be a symptom of lesions of the sensory cortex, it corresponds in distribution to the loss of the sense of position and passive movement.

The cortex at the lower end of the somesthetic area (tongue and face area) appears from the studies of Bornstein to be the area for taste. This area lies adjacent to the motor cortex governing the muscles of mastication.¹¹ It had been generally taught, but on doubtful evidence, that the center for taste lay close to that for smell, namely, in the region of the hippocampal gyrus. It appears from the results of the stimulation of the human cortex during operations that the taste area extends deeply into the fissure of Sylvius. An electrical stimulus applied to the cortex above the circular sulcus (surrounding the insula), or the surface of the insula itself causes a sensation of taste—a "terrific tight sensation of taste" as one patient expressed it (Penfield).

In a SUBCORTICAL LESION of the sensory pathway (i.e., from the thalamus to cortex) disturbances in the cortical faculties obviously must

¹¹ This area probably extends into the upper bank of the Sylvian (lateral) fissure, for the sensation has been aroused in patients by stimulation of the upper bank where it joins the insula (Penfield and Rasmussen).

occur, since many of the impulses from which these faculties are integrated will fail to reach their destination. The defects, however, are much cruder, they are in terms of the primary sensations, not in accordance with the three cortical faculties. There is not the uncertainty and inconstancy of response characteristic of lesions of the somesthetic area. The patient either feels a sensation or he does not, and gives the same answer each time a given stimulus is repeated. The loss of the individual sensations is often severe but when the particular stimulus is increased a response can usually be obtained.

Motor effects of a generalized type are produced by electrical stimulation of the posterior part of the superior parietal lobule (area 5, parietal adversive field). These are movements of the head and eyes to the opposite side. Stimulation of the angular gyrus causes conjugate deviation of the eyes to the opposite side.

Attacks of Jacksonian epilepsy, due to lesions of the sensory cortex, may be preceded by sensory auras—comprising pricking sensations, “pins and needles”, sensation of cold, etc.

THE TEMPORAL LOBE THE CENTERS FOR HEARING AND SMELL

The primary cortical center for hearing is situated in the transverse gyrus of Heschl lying in the floor of the lateral cerebral (Sylvian) fissure, and an adjoining small area of the superior temporal gyrus. Fibers from the medial geniculate body reach this *audiosensory* area (area 41) via the posterior limb of the internal capsule, they constitute the *auditory radiation* (p 1189). In the audiosensory area the fundamental auditory sensations—intensity, quality and pitch—are appreciated. The area is bilaterally represented.

Equilibratory sense is represented in the posterior part of the first temporal convolution. Stimulation of this region in conscious patients causes dizziness or nausea, a sense of swaying, falling, or of rotation. The temporal cortex also appears to be that part of the brain from whence pictures of past events can be recalled—that is, *memory*. Here too, probably, is where those fantastic patterns of garbled memories which we call dreams are woven.

As in the case of the somesthetic (p 1040) and visual senses, auditory sensations have a dual representation in each cerebral hemisphere—*auditory areas I and II*. Woolsey and Walzl

stimulated electrically the different levels of the exposed cochlea in cats and found a point to point projection on to the temporal cortex. In auditory area I, the apical turns of the cochlea were projected posteriorly, the basal turns anteriorly, whereas in the “secondary” area, which lies adjacent and ventral to the “primary” area, the different points of the cochlea showed the reverse distribution, apical turns anterior, basal turns posterior (fig 68.8). A large part of the cortex of the superior temporal gyrus lying outside these audiosensory areas is considered to be *audito-psychic* in

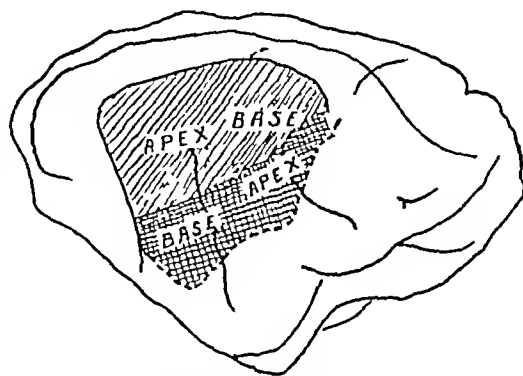


FIG 68.8 Right temporal cortex of cat showing the representation of the right (ipsilateral) cochlea. This is a composite figure from data of a number of experiments. Upper hatched region, primary auditory area, lower cross-hatched region, secondary auditory area. (After Woolsey and Walzl, redrawn with minor modifications.)

function. Herein the analysis and interpretation of auditory sensations, and their integration into more complex perceptions take place. The audito-psychic area is mainly unilateral, being on the left side in right-handed individuals and vice versa.

Fibers descend from the cortex of the temporal lobe to (a) the *medial geniculate* body and *inferior colliculus*, the former is therefore connected with the auditory area by both ascending and descending paths. (b) The *nuclei of the pons*. These fibers constitute the *temporopontine tract* which traverses the posterior limb of the internal capsule and the outer part of the base of the cerebral peduncle. Through these fibers and the pontocerebellar tract the temporal lobe is in communication with the cerebellum. The temporal lobe is connected also with the thalamus by both ascending and descending fibers.

The *CENTER FOR SMELL* is situated in the uncus and the anterior part of the hippocampal gyrus.

(pyriform area) In man, stimulation of the uncus, olfactory lobe or the region of the amygdaloid nucleus causes an olfactory sensation. Owing to the close relationship of these parts of the rhinencephalon to the temporal lobe, lesions of the latter (e.g., tumor, abscess, etc.) are not uncommonly associated with disturbances of the olfactory sense.

A LESION OF THE TEMPORAL LOBE may result in—

(a) *Aphasia* In a series of left temporal lobe lesions reported by Frazier and Rowe this disorder was present in 36 per cent. Others have reported higher percentages (e.g., Kolodny, 57 per cent).

(b) *Auditory disorders*, deafness, tinnitus (ringing or buzzing, etc.) or auditory hallucinations (e.g., hearing voices) may occur.

(c) *Olfactory and gustatory disorders*, impairment or loss of smell, or olfactory hallucinations, the subject imagining he smells some disagreeable odor. Taste may be defective or there may be disagreeable gustatory sensations.

(d) *Dreamy states* The subject experiences a sensation of unreality in which hazy memories of long past events are awakened. Such states are due to involvement of the region of the uncus and are sometimes on this account called uncinate attacks. In normal persons certain odors arouse a milder but somewhat similar sensation.

(e) *Disturbances of memory*

(f) *Hemianopia* (pp 1164–1167)

(g) *Epileptiform seizures* The convulsions may be ushered in by adverse movements (fig 682) and are frequently preceded by an auditory, olfactory or gustatory hallucination (aura), or by the dreamy state. The fit is sometimes precipitated by a sudden noise.

THE OCCIPITAL LOBE

The gray matter forming the walls of, and surrounding the calcarine fissure (on the medial aspect of the occipital lobe) constitutes the primary cortical center for vision—the *visuosensory area*. From the broad stripe of Gennari which can be seen with the naked eye this area is commonly known as the *area striata* or, following the numerical terminology, as area 17. It will be considered in more detail in chapter 76. Its histological features have already been touched upon (p 1031).

The *second visual area* wherein the visual sensations are interpreted and integrated into more complex perceptions is contiguous to the area

striata and lies on the lateral aspect of the occipital lobe (area 18).

Stimulation of the anterior part of the lateral surface of the occipital lobe causes conjugate deviation of the eyes to the opposite side (occipital eye field, area 19). In man, visual hallucinations, such as, flashes of light of different colors, or definite images have been evoked by the electrical stimulation of area 18 or 19.

THE PHYSIOLOGY OF SPEECH AND SOME OF ITS DISORDERS

The first stage in the development of speech is the association of certain sounds—(words)—with visual, tactile and other sensations aroused by objects in the external world. These associations are “stored” as memories. After definite meanings have been attached to certain words, pathways between the auditory area of the cortex and the motor area for the muscles of articulation become established, and the child attempts to formulate and pronounce the words which he has heard. This act of verbal expression involves the coordinated movements of a large group of respiratory, laryngeal, lingual, pharyngeal and labial muscles. Later, as the child is taught to read, auditory speech is associated with the visual symbols of speech, and finally, through an association between these and the motor area for the hand, the child learns to express his auditory and visual impressions by the written word.

APHASIA

This term is applied to those disorders of speech resulting from defects in the nervous mechanisms underlying the comprehension and use of symbols (words, numerals) for the formulation, transmission and reception of ideas. General intelligence may be little impaired. Yet aphasia is not simply a defect in the pronunciation of words as a result of the paralysis of the muscles of articulation. The innervation of the latter—motor area, corticobulbar fibers, cranial nuclei or peripheral nerves—is not necessarily affected. The defects in aphasia involve higher neural levels, they lie in the psychological sphere.

The faculty of speech is based upon a highly complex neural mechanism and being one possessed by the human brain alone cannot, of course, be investigated in animals, apart from the very limited opportunity afforded by exposure of the human brain during intracranial operations an

experimental approach to the study of speech and its defects is out of the question. In conscious patients electrical stimulation of the cortex of either hemisphere within the lower part of the precentral gyrus (lips, jaw and tongue areas), or in the upper part of the supplementary motor area on the mesial aspect of the hemisphere, causes the emission of a crude vocal sound like the cry of an infant or of an epileptic at the beginning of an attack. The vocal response elicited from the lip, jaw and tongue region never even remotely resembles a spoken word, but that evoked from the supplementary motor area is more complicated, and may have some slight resemblance to a word. Ablation of one or other of these areas causes only temporary speech defects. If the cortex in the lower part of the motor area of the dominant hemisphere (left in right-handed persons) is excised a temporary aphasia results which is probably due to some interference with the blood supply to area 44 (Broca's area). If while the patient is speaking, certain cortical areas are stimulated, speech is arrested, he cannot think of a word or words which he wished to use. Such speech arrest or induced aphasia has been used to map out area of cortex essential for normal speech. They are situated in the dominant hemisphere and are four in number: (1) *lower frontal*, area 44, (2) *upper frontal*, motor cortex anterior to the foot area on the mesial aspect of the hemisphere, (3) *parietal*, posterior to the lower part of the post-central gyrus, (4) *temporal*, posterior part of temporal lobe. Injury to one or other of these areas, with the possible exception of (2), causes persistent aphasia (see fig. 689).

Before giving Head's views on aphasia, a short account of previous ideas on the subject may be helpful to the reader.

The views of the neurologists of the 19th century had the merit of simplicity. These observers, to whom Head refers as the "diagram makers", conceived of the language faculty as built up of four separate components. Each of these, supposedly, was represented by a definite anatomically circumscribed area of the cortex, and could be affected independently of the others.

Two of these centers were *sensory* and two *motor*. All four were linked together by association tracts. In such a scheme, memories of spoken words were stored in the superior temporal convolution—the audiotopsychic center, the cortex in the region of the angular gyrus was the repository of visual word memories. These two receptive areas com-

prised what was referred to as Wernicke's zone. The pair of motor centers, i.e., those presiding over the coordinated movements concerned in vocalization and writing were called respectively the *glossokinesthetic* and *cheirokinesthetic* centers. The former was located in the posterior part of the left 3rd frontal convolution (area 44), the latter in the hind part of the left 2nd frontal convolution, in most instances on the left side. Diagrams were drawn confidently to show these four centers with their interconnections, and the type of aphasia which would result from the destruction of one or other component part of the neural mechanism. Broca (1861) believed that *motor aphasia* (see below) was the result of a lesion of the *glossokinesthetic* area (Broca's area), especially of the left side. Defects in the ability to write—*agraphia*—were held to be the result of the involvement of the *cheirokinesthetic* center. *Sensory* aphasias were classed as *auditory*—loss of the comprehension of audible speech—and *visual*—the inability to understand written or printed words. The former was held to be due to injury of the second temporal gyrus, the latter to injury of the angular gyrus. The most extreme proponents of this view even considered that every memory, auditory or visual, had its anatomical representation, so that a lesion limited to a small group of nerve cells would cause the loss of only those word memories for which they served as centers.

The different types of aphasia which were recognized may be briefly described.

(a) **MOTOR APHASIA (Broca's)** This term was applied to the type of speech defect in which the patient is almost speechless, but there is no paralysis of the muscles of articulation. Though unable to express his thoughts in words, he can understand what is said to him and can read. He is usually able to utter a few words of an ejaculatory nature, e.g., "oh my", "dear me", "damn", etc. Sometimes he is able to say the last words which he had spoken just before the onset of his illness, as in the oft cited case of the librarian whose only words were, "lists complete".

(b) **AGRAPHIA** This term indicated that the patient was unable to write though motor speech and the comprehension of written or spoken words were possible. The movements of the hand and arm for other acts were not impaired.

(c) **AUDITORY APHASIA OR "WORD DEAFNESS"** were the names given to those defects of the language faculty in which the subject, though able to hear, does not understand spoken words. He is

as a person listening to a strange language. The power of speech, writing and the comprehension of written or printed words are retained. He may be able to repeat words spoken to him, and instead of answering a question may simply repeat the questioner's words, this phenomenon is called *echolalia*.

(d) **VISUAL APHASIA OR "WORD BLINDNESS" (ALEXIA)** Vision is unaffected yet the recognition of written words or numerals and the appreciation of their meanings are defective or lost.

In 1906 these mechanical conceptions of the speech faculty and the production of aphasia were challenged by Pierre Marie. He claimed that

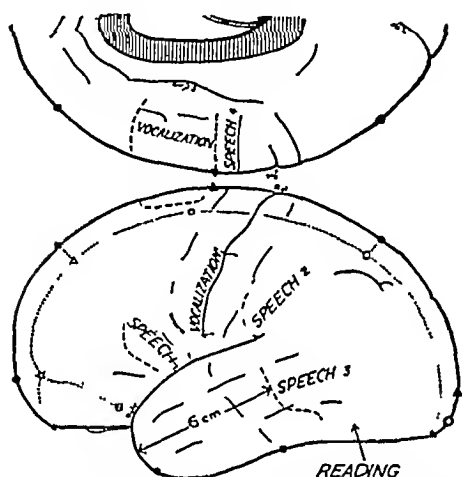


FIG 68.9 Showing speech areas in the left (dominant) cerebral cortex. (From Penfield and Jasper, *Epilepsy and the Functional Anatomy of the Human Brain*, courtesy of Dr Penfield.)

there was only one true type of aphasia—sensory aphasia—due to a lesion in Wernicke's zone, and associated with a lowering of those intellectual capacities upon which were based the use of symbols in the expression of ideas. Thus, not only was the comprehension of written or spoken words defective but the ability to solve problems in arithmetic, and to perform certain other mental tasks were impaired. He maintained that the classical motor aphasia of Broca was simply *anarthria* (see below) due to involvement of the motor cortex governing the muscles of articulation plus sensory aphasia, and that a pure motor aphasia, i.e., a defect of speech due to the loss of "motor images" did not exist. Marie denied the existence of isolated cortical centers governing the different aspects of the speech faculty, and attacked the belief that aphasia was the result of the destruction of images—motor, auditory or

visual—in such specific areas. He reexamined the brains of Broca's first two patients, which had been preserved in the Musée Dupuytren in Paris, and found no confirmation of the view of Broca that the posterior part of the third frontal convolution was necessarily involved in aphasia. In the first brain, the lesion was not confined to Broca's area but involved Wernicke's zone. The second specimen did not show a circumscribed lesion but, on the contrary, a generalized wasting, the posterior part of the third frontal convolution was not especially implicated.

The terms in the classification given above, e.g., motor aphasia, *agraphia*, "word deafness", etc., are still sometimes employed, but it is understood they refer to the outstanding features of a given case rather than that they denote clear-cut types.

Head's classification of the aphasias

Head's views, based upon an exhaustive study of patients suffering from gunshot wounds of the cortex, are also strongly opposed to the conception of the speech faculty being dependent upon circumscribed anatomical centers. He considers speech to be a highly integrated cortical process—a special aspect of intelligence—yet one which can suffer with little lowering of the general intellectual level. The different components of the speech faculty, he decides, cannot be separated from one another by disease. Nor, consequently, can the types of aphasia be classed as "motor" or "sensory," but in any type deficiencies on both the receptive and the executive side can be demonstrated. Head concluded that cortical representation of speech mechanisms was more diffuse than had been supposed and that strict localization was impossible. Aphasia is a state in which the power to use words and other symbols as instruments of thought and expression is affected or, as he expresses it, "aphasia is a defect in symbolic formation and expression." The more complicated or abstract the idea which must be understood or expressed, the greater is the difficulty. Thus an aphasic may be able to name *objects* correctly but fails to find the word for a more *abstract idea*, e.g., color. Shown a black object, for instance, and asked to name its color he fails to do so, yet indicates that he recognizes that it is black by saying "what you do for the dead." An aphasic soldier when shown a red object said, "what the staff wear."

Head, as a result of his investigation of aphasic patients, devised a series of six tests of graded

severely These tests are briefly as follows

(1) *Naming and recognition of common objects* Six objects, e.g., a pencil, key, knife, etc. The patient is asked to name each object as it is pointed out to him. Next he is asked to point to each object as its name is called out. He is then given cards upon each of which the name of one of the objects is written. He is asked to indicate the object named.

(2) *Naming and recognition of colors* This test is carried out in a manner similar to that described for test (1) except that eight strips of differently colored silk are substituted for the six objects.

(3) *Man, cat and dog test* is designed to investigate the powers of reading and writing in their most elementary forms. The printed words "man," "cat" and "dog" are employed. The subject is asked to read these words, to write them from dictation, to copy them or to repeat them after hearing them spoken. Pictures of a man, a cat and a dog are also shown and the patient asked to write or to speak their names.

(4) *The clock tests* The patient is requested to set the hands of a clock in the same positions as those of a similar one set by the observer. He is then told to set the clock from verbal or printed commands. Again, he is asked to state the time aloud or by writing, of a clock set by the observer.

(5) *Coin bowl tests* Pennies are placed one in front of each of four bowls. The patient is asked verbally and in writing to place a coin in one or other of the bowls according to their number in the row. He is then asked to give an order himself and to carry it out according to his own words.

(6) *The hand, eye and ear tests* The patient is requested to repeat the movements of the observer which consist in touching an eye or an ear with one or other hand. When this is done correctly the patient's hand which moves is of course diagonally opposite to the hand of the observer. A much easier form of this test is the imitation of the observer's movement as reflected in a mirror. This simply requires matching without calculation.

The aphasic frequently fails to recognize that when the one hand is brought to his contralateral eye or ear the hand crosses the face. A further part of the test is to ask him to imitate the positions shown in pictorial form upon cards or to carry out the movements from printed and verbal instructions. Finally, he is asked to write down movements made by the observer.

Employing these tests Head divided aphasia into four types, as follows

(1) **VERBAL DEFECTS** The outstanding feature is a defect in the utterance of individual words of all kinds. The power to express an idea in words is practically lost. The patient, however, is not entirely speechless but can usually utter a few

monosyllables, "yes" or "no", etc., or ejaculations and emotional expressions, such as, "damn", "oh dear me". When the disorder is less severe, the words are mispronounced but sentences are correctly constructed. For example, one patient said that he had trouble with "tenical terms" (technical terms) and that he "had *diffulty* in remembering what you do with a skull, *tri-tre-trispine*" (trepan). Another spoke of "*claration* of war by the *Ollies*" (declaration. Allies). Another would say "pyramerad" (pyramid), "sis-siors" (scissors) and "oboid" (ovoid). Such patients read with difficulty and writing is very defective or impossible. They usually understand printed or oral commands. This form of aphasia resembles the classical motor aphasia described on page 1046.

(2) **SYNTACTICAL DEFECTS** (agrammatism, jargon dysphasia). The patient is voluble but speaks a jargon in which, though the individual words may be fairly accurately pronounced, they are strung into short phrases or badly constructed sentences without articles, prepositions or conjunctions. The ability to read aloud is impaired, and curiously enough such a patient, though he can write a well constructed letter, may be quite unable to read it coherently. Such a one when asked the contents of a letter which he had just written replied, "I can't, I know, I suppose in time, not now, funny thing, why". In other instances the words themselves are often slurred over, mutilated and may be unrecognizable. Speech sometimes resembles "baby talk". Thus one patient when asked what his right arm felt like replied "Tiffrent from uffer um" (different from other arm). The understanding of ordinary conversation is defective.

(3) **NOMINAL (NAMING) DEFECTS** In this form of speech disorder the patient has difficulty in finding the right word to express his meaning or in naming a well known object. He will often employ a descriptive phrase in substitution for the word which he cannot recall. For example, a painter when asked to name a series of colors could not say "violet" but instead explained that "it was made with black, red and a bit of blue". Another when asked to tell the time from a clock which had both hands at 12 replied "That is when you eat." These patients can draw from a model either directly or from memory, after it has been shown and then removed, but are usually unable to draw from imagination. They write a coherent letter with difficulty, usually fail to carry out sim-

ple arithmetical exercises and confuse the values of coins

(4) SEMANTIC¹² DEFECTS A patient suffering from this type experiences little difficulty in articulate speech, can name objects, understands individual words and some sentences, but the general meaning of what he hears escapes him. He often fails to follow his own utterances to an intelligent conclusion, his sentences tailing off as though he had forgotten what he had started out to say. When shown a picture he picks out the details but fails to grasp the meaning which it conveys to others. Such a patient therefore misses the point of a joke whether this is printed, told to him, or is in pictorial form. He fails to comprehend the significance of much that he sees and hears. There is no impairment in the pronunciation of words and, though speech tends to be in short jerky sentences, syntax and intonation are not disturbed.

Head, though he discards the conception that the neural basis of speech consists of strictly localized anatomical "centers" wherein resides *exclusively* one or other of the speech functions—auditory, visual or motor—believes that regions exist in the cortex "where the progress of some mode of action can be reinforced, deviated or inhibited." These regions constitute foci of integration—convergence points for association paths. Destruction of one or other of such foci or "knots" of association paths will depress *as a whole* the psychological processes underlying speech. The speech faculty is disabled, certain faculties are lost, while others are retained. Yet he points out that it is not logical to conclude that the abilities which remain and those which have been abolished constitute essentially separate and distinct functions from which the normal processes of speech have been synthesized, or that they are represented in specific circumscribed areas. To make a rather crude comparison—a person who has injured his foot, knee or hip, adjusts his locomotor apparatus as best he can. He hops or limps, yet it can not be argued that the hopping or the limping motion which he employs is simply one of the component movements employed in the normal act of walking and which the injury has left intact. Nevertheless, the form which the disability assumes is undoubtedly influenced by the site of the injury. The ambulatory abnormality, for example, which results from an injury to the foot is different from that resulting from

injury to the knee or hip. So too the nature of the speech disability is influenced by the particular region of the cortex involved. Thus if the lesion is in the neighborhood of the lower part of the precentral and postcentral convolutions of the dominant hemisphere the speech defect tends to be of the *verbal* type. In injury to the temporal lobe the speech defect tends to be of the *syntactical* type. In a lesion in the region of the angular gyrus of the dominant hemisphere the patient has difficulty, particularly, in finding names for things (*nominal defect*), damage to the cortex in the region of the supramarginal gyrus results in a *semantic defect*.

Anarthria or dysarthria is loss or difficulty of speech due to paresis, paralysis or ataxia of the muscles concerned in articulation. There is no impairment of the psychical aspects of speech, i.e., "internal speech" is unaffected, there is no difficulty in the comprehension of spoken or written speech. Other functions, e.g., swallowing, which are dependent upon the same groups of muscles as those used in speech, are also frequently affected. The condition may result from a lesion in the internal capsule or corpus striatum, bulbar nuclei or peripheral fibers, or from disease of the muscles themselves. Since the muscular mechanism of speech is innervated from both sides of the brain unilateral lesions are not followed by permanent anarthria. A lesion of the cerebellum or of its connections may also cause disordered control of the muscles of articulation (p. 1083).

APRAXIA AND AGNOSIA

Apraxia (unable to act) is the inability to perform purposeful movements at will, either at command or in imitation, though the muscles normally engaged in the act are not paralysed. It is allied to aphasia, which might be called apraxia of the speech faculty. Apraxia may be sensory or motor. In the former, the patient does not recognize the significance of an object (visual agnosia), and therefore cannot put it to its proper use, this is simply visual agnosia. When, for example, he is given a pencil he may, upon a request to use it, attempt to clean his teeth with it or smoke it like a pipe. In motor apraxia the patient has no conception, or a very defective one, of the pattern of muscular movement required to perform a purposeful act. For example, apraxia of the tongue is frequently seen in hemiplegic patients. The tongue cannot be protruded upon request, but a moment later the patient may without thought lick his lips. When given an object, and asked to use it, though he recognizes it and knows its use, he cannot form the "motor picture" required to execute the act but manipulates the object in an awkward aimless manner. The defect is evidently in the physical sphere.

¹² Semaine = to signify

and not due to disease of the cells of the motor area or of the pyramidal fibers. It is thought to be due to the interruption of association tracts connecting the precentral gyrus with higher psychical regions of the cortex where impressions of the movements of muscles are received, synthesized and stored as kinesthetic memories. This higher ideational area probably lies in the region of the left supramarginal gyrus in right handed persons. A lesion of this region may cause bilateral apraxia, one confined to the anterior part of the corpus callosum is likely to interrupt fibers passing from the left hemisphere to the right precentral convolution and so cause apraxia of the left side.

Agnosia (not knowing) is a defect of a higher level of consciousness than the mere inability to perceive tactile, visual, auditory or other forms of sensation, it results rather from the failure to interpret sensory impressions which enable an object, sound, symbol, etc., to be recognized and have meaning. A patient suffering from auditory agnosia, for example, cannot appreciate music or the meaning of other sounds. "Word blindness" and "word deafness" are forms of auditory and visual agnosia, respectively. The subject of visual agnosia is unable to name an object, not that he is aphasic in the true sense, but simply because the object is quite strange to him. When shown an object and asked to use it he behaves quite differently from the patient with motor apraxia who recognizes it but is unable to perform the necessary movement.

Astereognosis is a form in which though sensations of touch and muscle sense are retained, the patient cannot recognize an object placed in his hand if his eyes are closed. Visual agnosia is seen in lesions of the occipital lobe of the dominant hemisphere, auditory agnosia in injury to the temporal cortex, and astereognosis in lesions of the parietal lobe posterior to the postcentral gyrus.

Agraphia or dysgraphia, the inability to write or difficulty in writing, is usually associated with visual agnosia in so far as the recognition of written words is concerned, i.e., word blindness.

EPILEPSY

Epilepsy is a condition characterized by recurring attacks or "fits", which when exhibited in typical and severe form, consist of abrupt loss of consciousness and generalized convulsions. Two stages of the attack or seizure are recognized. In the first or *tonic stage* the muscles contract tonically, the spasms often twisting the facial features and holding the head and limbs in distorted positions. The arms are most commonly flexed and the lower limbs rigidly extended. After a few seconds the tonic spasm gives place to jerking movements, often violent, of the limbs, face and muscles of mastication. This is spoken of as the

clonic stage. Either during this stage or in the tonic stage the tongue may be bitten. Before the onset of the convulsion a large proportion of epileptics receive a warning in the form of a sensation or hallucination, the character of which varies in individual cases. The warning sensation or *aura*, as it is called, may be auditory, e.g., voices, music, etc., visual, e.g., flashes of light, sparks, etc., olfactory, gustatory, cutaneous, visceral or vasomotor, equilibratory, or anesthetic, i.e., a sensation of movement of some part of the body. Turning of the head and trunk to one side and deviation of the eyes are commonly observed. The patient sometimes utters a cry or scream—the epileptic cry—just before consciousness is lost. Attacks of sudden loss of consciousness, usually of brief duration, may occur without convulsions, such minor seizures are referred to as *petit mal* to distinguish them from the major attack or *grand mal*.¹³ A third form known as *psychic* or *psychomotor* epilepsy is marked by automatic movements, such as smacking of the lips, chewing, together with a clouded, "dreamy" feeling of unreality, or of having seen or heard before some sight or auditory impression which is actually happening at the moment ("déjà vu" or "déjà entendu" phenomenon, respectively). Or there may be a confused mental state persisting for a minute or two, or for a longer period. During this time the patient may perform automatic acts of which he is quite unaware and does not remember. Unlike *grand mal*, generalized convulsions do not occur nor does the subject fall to the ground. In this type the primary discharge is thought to be in the cortex of the temporal lobe, focal epilepsy showing such features occurs in lesions involving the region of the cortex. (See also diencephalic autonomic epilepsy, chapter 57.) After the convulsion the subject remains for a time in a stupor or may, especially after milder attacks, perform automatic acts, of which he has no recollection after regaining consciousness. This *automatism* occurs also in attacks of *petit mal*. Most commonly it consists merely in resuming some act or other upon which the subject was engaged before the onset of the seizure. Sometimes a number of convulsive seizures occur in rapid succession, the patient failing to regain consciousness in the intervals between them. This very serious condi-

¹³ There is no essential difference between *petit mal* and *grand mal*. They are but degrees of intensity of the disturbance.

tion is called the *status epilepticus*. Electroencephalograms in epilepsy are shown in figure 68 11

PATHOGENESIS

With respect to origin two forms of epilepsy are recognized—*Jacksonian* or *focal* and *idiopathic* or *cryptogenic*

(a) *Jacksonian epilepsy* In this type the seizure is due primarily to a gross lesion localized in some part of the cerebral cortex. For example, a tumor, a foreign body, a cerebral birth injury or a depressed fracture of the skull may, by stimulating the cortical tissue in its neighborhood, precipitate the epileptic fit. A seizure arising from a definite cause of this nature is spoken of as Jacksonian epilepsy after Hughlings Jackson, the English neurologist of the last century who first described it, or as focal or symptomatic epilepsy. The fit commences in the muscles governed by the irritated area of the cortex.

Twitchings of the fingers, great toe or lips, or more pronounced clonic movements of these parts, are usually first observed, but the disturbance spreads (*Jacksonian march*) to involve other regions, and finally a general convulsion develops. In the "march" of the convulsion the order in which the different muscles are involved can frequently be seen to correspond to that of their cortical representation. When the lesion is in the frontal, parietal or occipital lobes, and involves the eye fields or the adversive fields, the attack is often ushered in by eye movements or turning of the head and trunk and deviation of the eyes to the opposite side.

(b) *Idiopathic or cryptogenic epilepsy* Epilepsy which cannot be explained by the presence of any organic lesion of the brain is called idiopathic or cryptogenic. Though many theories have been advanced, the cause of this type of the disease remains obscure. There is a tendency today to look upon epilepsy as a symptom, or rather a group of symptoms common to several rather than to a single primary pathological state. Convulsions, whose features are indistinguishable from those of the epileptic seizure, occur in a number of conditions. In animals convulsions may be produced by injections of absinthe or caffeine. Hypocalcemia (ch 60), hypoglycemia (ch 49), cerebral edema or anemia and other states are accompanied by generalized convulsions of an epileptiform character.

With regard to the neural mechanism through which the fits are produced, some authorities have thought that the convulsions are the result of

increased excitability of the cortex and therefore comparable in their mode of production to those of the Jacksonian type, or to those produced by experimental stimulation of the cortex. Others view the convulsions as a *release phenomenon* due to the inhibition of cortical areas which normally exert a controlling influence upon lower motor centers, e.g., weakening of suppressor action (p 1039). Finally, there is the view of Penfield and his colleagues that the primary disturbance in idiopathic epilepsy has its origin in a subcortical mechanism.

The close correspondence between the convulsive seizure as seen in focal epilepsy or to artificial stimulation of the cortex, and the fits of idiopathic epilepsy appears to support the first alternative as a cause of the generalized seizure known as grand mal. However, the *primary* disturbance responsible for the cortical discharge in idiopathic epilepsy probably lies outside the cortex. Evidence for a subcortical origin of petit mal attacks is provided by electroencephalographic studies. During such an attack, electrical potentials of the wave-spike type (fig 68 11)—appear synchronously over prefrontal areas as well as from other widely separated cortical areas. This suggests that the cortical discharges are initiated in a common subcortical region (centrencephalic system) connected with extensive areas of the cortex. Jasper and Drooglever-Fortuyn, experimenting with cats produced electrical potentials from wide spread areas of the cortex synchronously in both hemispheres by rhythmical stimulation of a small area (2 mm in diameter) in the anterior part of the massa intermedia of the thalamus. The cortical potentials were of the wave-spike form characteristic of petit mal.

Moreover the general convulsions which occur in focal epilepsy are thought to be due to the discharge of impulses from the cortical focus into the subcortical mechanism and causes unconsciousness. The subcortical structures then fire back and cause widespread excitation of the cortex of both hemispheres.

The nature of the underlying bodily state responsible for the convulsive seizures of idiopathic epilepsy are unknown. There are a number of drugs, e.g., metrazol, which are capable of inducing convulsions, and several others, e.g., phenobarbital and tridione, which are anticonvulsive, but they have not led to an understanding of the pathogenesis of the convulsive seizures of

epilepsy Susceptibility to an epileptic seizure is influenced by several physiological and biochemical factors which are given in table 95, but they do not bear a specific relationship to the cause of idiopathic epilepsy, but merely affect the tendency to convulsive seizures of any kind

A predisposition to epilepsy is an inherited characteristic which is associated with and can be revealed by the electroencephalogram Dysrhythmia consisting principally in an abnormal slowness of the electrical waves is found in the relatives of epileptics In a study of some 200 parents of epileptic children, only 5 per cent gave normal electroencephalograms Dysrhythmia was found 6 times more frequently in the parents of the epileptics than in the general population, and in 35 per cent both parents showed abnormal waves Predisposition is thought to be inherited as a Mendelian dominant Yet, of the children in most family groups with an epileptic ancestry only about 3 per cent actually develop epilepsy

The view that the actual seizure was initiated by cerebral ischemia resulting from a spasm of the cortical vessels has not been sustained The blood flow through the brain of epileptics and the oxygen consumption have been found to be no different from that of non-epileptics In focal epilepsy, however, local ischemia may play a part, for the capillaries are reduced in number in the tissue adjacent to the injured area Biochemical studies of such epileptogenic areas have not however revealed any important abnormality The pH of the extracellular fluid shows no change from the normal, or a slight rise

The induction of a dehydrated state of acidosis (by means of a ketogenic diet) has been employed to reduce the susceptibility of epileptics to seizures, and pitressin combined with an increase in the water intake is used as a test (*water-pitressin test*) in diagnosis But in only about 30 per cent of epileptics is a seizure precipitated, so a negative result does not exclude epilepsy

TABLE 95

	CONDITIONS WHICH MAY TEND TO	
	Prevent seizures	Precipitate seizures
Oxygen	Rich supply	Poor supply
Acid base equilibrium	Acidosis by means of fasting, fat diet Ingestion of acids or acid forming salts Breathing high CO ₂	Alkalosis by means of ingestion of alkali Hyperpnea—"blowing off" CO ₂
Water balance	Dehydration	Edema
Intracranial pressure	Decreased	Increased
Serum calcium	Increase	Decrease (hypocalcemia)
Blood glucose	Increase	Decrease (hypoglycemia)
Acetylcholine	—	Increase
Cholinesterase	—	Decrease
Body temperature	Increase	Decrease

HEADACHE

Headache is one of the commonest of symptoms, and occurs in a great variety of diseased states, e g, arterial hypertension, chronic nephritis, eye strain, chronic disease of the paranasal sinuses, brain tumor, constipation, etc.

Headache may also be due to neuralgia of the scalp, or from osteitis or perostitis of the cranium (syphilis, Paget's disease, etc.) It may sometimes be due to pain (ch 44) referred from a thoracic or abdominal viscus to the superficial tissues of the head supplied by the trigeminal, which is related segmentally to the vagus The temporary but severe pain felt in the frontal region by a normal person after swallowing ice-cream is an example of a referred pain This pain is most probably due to the stimulation of afferent vagal endings in the lower part of the esophagus, or in the stomach, and referred to the forehead through central connections with the sensory nucleus of the trigeminal nerve

The two principle causes of the more common types of headache are either vascular in nature (stimulation of sensory nerves of the intracranial vessels) or a sustained reflex contraction of the posterior neck muscles inserted into the scalp, especially of those terminating in the vertex and occipital regions This latter reflex mechanism is mainly responsible for the headache of eyestrain—refractive errors, or unbalanced action of the

extra-ocular muscles, excessive and prolonged accommodation—increased intra-ocular pressure also may contribute to the pain. The headache associated with an intracranial tumor, and with hypertension, are of vascular origin. The arteries and smaller veins on the convexity of the human hemispheres have been shown by neurosurgeons to be insensitive, but the larger arteries in the Sylvian (lateral) fissure and at the base of the brain, e.g., circle of Willis, vertebral, basilar, the larger of cerebral vessels of the dura, the branches of the anterior and posterior meningeals, as well as the arteries of the scalp, contain pain fibers which are stimulated by any stretching force. The larger veins where they empty into the dural sinuses are also extremely sensitive. The brain substance itself is painless. The headache of intracranial tumor or abscess is not due primarily to the high intracranial pressure itself, but is the result of the displacement of the brain, local mechanical effect of the tumor, or to blockage of some part of the ventricular system. In this way traction or distortion of the vessels with consequent stretching of the sensitive vascular nerve endings, and, in some cases, of cranial nerves, results. In a large proportion of cases of headache due to brain tumor, the intracranial pressure is not elevated (Wolff and Wolff), and when headache exists, lowering the pressure within the skull does not always relieve it. Raising the intracranial pressure in a normal person or one with an intracranial growth does not necessarily cause headache. On the contrary, lowering the intracranial pressure as by drainage of cerebrospinal fluid causes severe headache. A normal subject experiences headache when about 20 cc of fluid are withdrawn by lumbar puncture, and he is raised into the erect position. In the horizontal position, or if the fluid removed is replaced by an equal quantity of saline, the severity of the head pain is reduced.

Headache, presumably of vascular origin, can be produced in any normal person, as was shown by Wolff and his associates, by a sudden sharp rotary movement of the head. When a subject with headache due, apparently, to overdistension of the cerebral vessels was centrifuged, so as to drive blood out of the vessels, the pain was relieved. The intense headache caused by an injection of *histamine* appears to be due to cerebral vasodilatation and distension of the vessels, it is relieved by increasing the cerebrospinal fluid pressure, and thus providing a circumferential

support for the vessels, and protecting the pain endings from stretch. The headaches of fever and toxic states are similar in their origins to histamine headache, they, too, are ameliorated by raising the cerebrospinal fluid pressure.

Hypertensive headache is not due essentially to the high blood pressure (Wolff and Wolff), but rather to the amplitude of the movement of the arterial wall at each pulse beat which varies with the contractile state of the vascular wall itself. When the latter is relaxed and *resilient*, a given distending force will exert a greater degree of stretch upon the pain endings than when the walls are firmer and do not so readily permit stretching of the nerve terminals. The headache of arterial hypertension therefore varies with the amplitude of the movement of the arterial wall and is relieved by pressure upon the common carotid which reduces the pulse amplitude. Though the headache is not directly related to the height of the blood pressure, (since it may occur with equal intensity when the pressure is high, moderate or relatively low), with any given degree of relaxation of the vascular wall, the pain will of course be more intense when the pressure is high than when low, because a greater excursion of the vascular wall will occur in the former instance. Ergotamine while it tends to increase the blood pressure rather than to reduce it diminishes the amplitude of the arterial pulsations and as a consequence has a moderating effect upon the intensity of the headache.

The headache associated with constipation is, however, apparently of reflex rather than of toxic origin. It is often relieved very quickly by a movement of the bowels, a fact, as pointed out by Alvarez, which seems to preclude the possibility that it is the result of the absorption of toxins from the intestine.

The site of the headache in brain tumor may be of considerable value as a localizing symptom. The pain overlies the tumor or is on the same side in about 60 per cent of cases. In tumors of a cerebral hemisphere (supratentorial tumors) the pain is usually in the front of the head, the pain impulses being transmitted through branches of the 5th nerve, whereas in those of the posterior fossa the pain is felt in the occipital region, the impulses travelling by branches of the 9th and 10th cranial and upper cervical nerves.

MIGRAINE (Synonyms, *megrim*, *hemicrania*, *sick headache*.) This is a type of headache whose special features entitle it to be placed in a class by

itself. The headache is periodic, severe and often accompanied by nausea and vomiting. In many instances some type of sensory disturbance (aura) ushers in the attack. This frequently takes the form of scintillating colored lights or the so-called fortification figures, that is, zig-zag luminous bands which are suggestive of the walls of a turret. The visual hallucinations have a homonymous distribution, that is, they occur in the right or the left halves of the visual fields (p. 1168). Temporary hemianopia may follow or blindness of the central part of the retina may accompany the visual sensations, which are then toward the periphery of the field of vision. Cutaneous, auditory or gustatory auras occasionally occur. The headache is localized at first but soon spreads to involve the entire half of the head, the pain is then on the side opposite to that of the hallucinations. Not unusually, however, the headache is bilateral, or is unilateral to start with, and later spreads to involve the opposite side of the head. In most cases of migraine, the pain has its origin in extracranial vessels, the branches of the external carotid, such as, the temporal or occipital artery being most commonly involved. When the pain originates intracranially the anterior and middle meningeal arteries are probably implicated.

The mechanism underlying the migrainous headache is not universally agreed upon. It is thought by some (Wolff and Wolff) to have a vascular origin, similar in nature to that causing hypertensive headache, namely a relaxed state of the vascular wall, and a resulting pulsatile movement of greater than usual amplitude. The pain varies in intensity with the amplitude of the arterial pulsations, and is reduced in intensity by compression of the common carotid or by ergotamine which, as mentioned above, reduces the pulse amplitude. In other cases the headache of migraine appears to be due to cortical vasodilatation (like the histamine headache) which follows upon a period of vasoconstriction. The ischemia of the cortex during the vasoconstrictor phase would explain the visual hallucinations mentioned above, which an increase in pulse amplitude is unable to do, though the phase of vasodilatation might quite well be associated with a relaxed state of the larger cerebral vessels. It has been suggested that migraine may sometimes be of allergic origin, dilatation of the cortical vessels being the result of the local liberation of histamine (H-substance). The relatively high incidence of eosinophilia in migraine lends support to the idea of an allergic reaction.

The electro-encephalogram (EEG)

In 1929 Berger reported his discovery that changes in electrical potential could be recorded from the head of the human subject by means of pad electrodes applied to the scalp or needle electrodes placed in contact with the periosteum of the



FIG 68 10 Normal electroencephalogram taken from the occipital region. O and C refer, respectively, to open and closed eyes. (After Adrian and Matthews.)

skull. These brain potentials were later studied by Adrian and Matthews. In normal subjects three wave frequencies may be recorded, the *alpha*, *beta* and *delta* rhythms.¹⁴ The alpha rhythm consists of rhythmical oscillations in electrical potential occurring at the rate of 10 per second (fig. 68 10). The waves have a voltage of about 50 microvolts on the average. The beta rhythm has a frequency of 15 to 60 per second, and the waves are of lower voltage (5-10 microvolts), the frequency of the delta waves is from 1 to 5 per second and the voltage is relatively high (20-200 microvolts). No precise cytoarchitectural area can be said to have a characteristic rhythm as was once supposed, but differences are found between rather large areas of the cortex. Thus the alpha rhythm dominates records from the occipital region or from the posterior parts of the parietal and temporal lobes. The sensorimotor region emits waves with a frequency of from 20 to 25 per second, the anterior frontal areas give out waves at the rate of from 8 to 5 per second.

The alpha rhythm occurs in the *inattentive* brain, as in drowsiness or light sleep, narcosis or when the eyes are closed, it is abolished by visual and other types of stimulus or by mental effort (e.g., mathematical calculation). It therefore disappears when the eyes are open. The waves disappear even if the subject opens his eyes in the dark and tries to see. It is apparent therefore that it is the attention rather than the visual stimulus itself that abolishes the alpha rhythm when the eyes are opened. On the other hand if the visual field is uniform, that is without pattern, or glasses are worn which blur the visual image so that it has no meaning, the rhythm is not abolished. An attempt to discern any detail causes the alpha wave to immediately disappear. A visual field that flickers causes the waves to assume the same rate

¹⁴ A faster rhythm (*gamma*) appears in rare instances

as that of the flicker. The delta waves can be recorded very rarely from a normal adult while awake, but appear normally during deep sleep or during the waking hours in early childhood. Generally speaking, their presence in an adult, except during sleep, indicates some pathological process in the brain—tumor, epilepsy, raised intracranial pressure, mental deficiency or depression of consciousness by toxic or other factors. When present they tend to displace the alpha rhythm. Neither the beta nor the delta waves are affected by opening or closing the eyes.

or reduces the voltage of the waves in the cortex and greatly alters their form.

It is now very generally held that while the cortex is not incapable of a rhythmical electrical activity of its own which can be influenced and altered in various ways through the usual afferent channels, normally, the "resting rhythms" as ordinarily recorded are mainly dependent on, and maintained by the activity of neurons in the reticular formation of the brain stem and structures of the diencephalon (thalamus, hypothalamus and subthalamus)—that is, largely those parts of

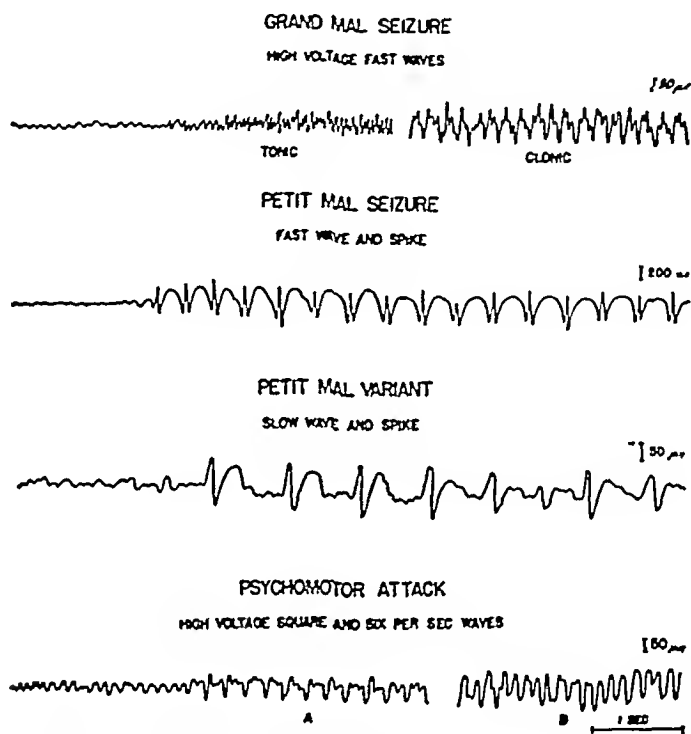


FIG 68.11 Showing electroencephalograms in three kinds of epilepsy (After Gibbs and associates)

There has been a difference of opinion as to the origin of the potential changes as recorded in the EEG. Whether they are inherent in the cortex, or are dependent upon subcortical structures has been argued. The latter alternative has received support from an experiment of Burns who isolated a slab of unanesthetized cortex completely from the rest of the brain except for its blood supply which remained intact. The isolated tissue responded when stimulated, but all spontaneous electrical activity was abolished. If a narrow bridge of nervous tissue was left connecting the slab of cortex with the rest of the brain spontaneous electrical waves crossed the bridge and excited the otherwise isolated cortex. Division of thalamo-cortical connections by undercutting also abolishes,

the upper brain stem which are now known as the centrencephalic system (p. 1039).

The EEG is affected profoundly by certain general states. *Hypoxia* causes at first a moderate slowing of the rhythm, but as the oxygen lack becomes more severe large delta waves appear, with persistent and severe hypoxia the amplitude of the waves declines and may reach almost the vanishing point. *Hypoglycemia* causes an effect somewhat similar to that of hypoxia, when the blood sugar falls below about 60 mg per cent, delta waves (1–3 per second) are seen, though the alpha rhythm is not abolished entirely. The effects of hypoxia and hypoglycemia enhance one another, their summed effects being much greater than the effect of each alone. *Hyperglycemia* has

little effect upon the cortical potentials. Increase in carbon dioxide tension increases the rate, but reduces the amplitude of the waves. Blowing off of CO_2 , as by overbreathing, has the opposite effect.

Abnormalities of the electro-encephalogram in brain tumor, especially of so-called silent areas of the brain, and in epilepsy, are of definite diagnostic value. In cerebral tumor or brain abscess, as shown by Walter and by Case, the functionally depressed brain tissue surrounding the lesion gives out waves of slow rhythm (delta waves) which, combined with a loss or diminution in the alpha rhythm over the occipital region may be of considerable aid in localization.

Epileptic seizures are characterized by pronounced departures from the normal rhythm. In *petit mal* attacks, large slow waves appear about a second before the attack is clinically manifest, and displace the previous rhythm. Each large wave is followed by a sharp spike deflection (fig. 68 11). In the tonic stage a *grand mal* seizure, waves

of relatively high frequency (10–30 per sec.) and of low voltage appear, but as the attack progresses into the clonic phase these fast waves give place to slower and large waves which continue into the stage of stupor following the seizure. Delta waves may be a prominent feature of the electro-encephalogram of epileptics between seizures (fig. 68 11).

In normal sleep the pattern of the electrical potentials recorded from the brain varies with the depth of unconsciousness. During light sleep delta waves make their appearance while the alpha waves superimposed upon the slower rhythm of the latter persist. In deep sleep the alpha rhythm disappears being replaced by delta waves, or in some instances by a faster rhythm with a frequency of about 14 per second. During the light sleep before awakening the record shows only an odd slow wave. As consciousness returns the tracing consists of an intermittent alpha rhythm which becomes continuous upon waking.

CHAPTER 69

CONDITIONED REFLEXES SLEEP

DEFINITIONS

The ordinary reflex with which we are all familiar is an inherited characteristic of the species and is not dependent upon previous experience. Its pathways are established at birth. This type Pavlov termed an *inborn* or *unconditioned reflex*. For example, food placed in the mouth of a newborn puppy evokes a secretion of saliva. The reaction depends solely upon the stimulation of receptors (taste, touch, etc.) in the mouth, the afferent and efferent nerves and the salivary centers in the medulla oblongata. When, on the other hand, the young animal sees or smells a piece of meat for the *first* time no secretion of saliva results. Yet if an animal who has eaten meat on previous occasions sees or smells a morsel, a profuse secretion of saliva occurs. This reaction, which depends upon previous experience, Pavlov has termed a *conditioned* or *acquired reflex*. Its pathways are not fully established at birth but are developed by training. Quite evidently the reaction of the older animal is the result of an association established in the past between the stimulus applied to the receptors of the mouth, and the appearance or smell of the food, i.e., a visual or olfactory stimulus. The former is called the *unconditioned*, the latter the *conditioned stimulus*. Not only the qualities of the food itself but changes in the environment extraneous to the food, if they occur in association with feeding, can serve as conditioned stimuli.

Conditioned reflexes may be either of an excitatory character, e.g., the secretion of saliva, or have an inhibitory action. The former are termed *positive* or *excitatory*, the latter *negative* or *inhibitory*.

POSITIVE OR EXCITATORY CONDITIONED REFLEXES

In the great majority of the experiments performed by Pavlov and his school, the secretion of saliva was chosen as the indicator of the conditioned response. In order to follow the secretory reaction with precision the opening of the parotid or submaxillary duct was transplanted to the cheek or chin, respectively. The saliva is collected by means of a special apparatus consisting of a funnel sealed over the duct opening and leading into a system of tubes. The secretion is measured in drops by means of an electrical recorder. The

animal is held in a stand by means of straps and occupies a sound-proof chamber separate from that of the experimenter (fig. 69 1).

The reflex is established in the following way. While the animal is being fed (unconditioned stimulus) a stimulus, e.g., a flash of light, which is quite alien to the food itself (conditioned stimulus) is applied. After this association of the two stimuli has been repeated a number of times, the flash of light alone (i.e., food is withheld) evokes a secretion of saliva. This is called a *conditioned alimentary reflex*. Motor reactions, e.g., movements of the lips and jaws, snapping, whining or barking, and movements of the limbs, accompany the salivary secretion and constitute an integral part of the reflex. The number of repetitions of the experimental procedure, or "lessons", necessary to establish the reflex varies in different animals and in experiments of different types. Many types of conditioned stimulus (visual, auditory, olfactory and cutaneous) have been employed by Pavlov and his associates.¹ In establishing the reflex the conditioned stimulus must precede the unconditioned stimulus and, except in the case of secondary and trace reflexes, to be presently described, must overlap it for at least a brief period. If the conditioned stimulus *follows* the actual feeding it is quite ineffective, i.e., it will not evoke a reflex when subsequently applied alone. After the period of training, in order to demonstrate the conditioned response satisfactorily, the animal should be alert, not drowsy, and preferably hungry, furthermore it should not be distracted by some extraneous stimulus, e.g., a strange sound or light, which sets up unconditioned motor reflexes.

OTHER TYPES OF POSITIVE CONDITIONED REFLEXES

CONDITIONED DEFENCE REFLEXES Acid injected into an animal's mouth (unconditioned

¹ Among these are the following,—the sound of a metronome, horn, bell, buzzer, tuning fork, organ pipe or of bubbling water, variously shaped objects, lights, figures or rotating discs, thermal, tactile and painful cutaneous stimuli, the odor of such chemicals as amyl acetate or vanillin. The cessation of a previously continuous stimulus, e.g., a buzzer, the rapid change in the intensity of a stimulus, or even the change in the rate of a rhythmical sound, e.g., a beating metronome, may serve as a conditioned stimulus (see also p. 1061).

stimulus) causes a profuse secretion of saliva which washes away the offending material. A conditioned reflex to acid is readily established by a series of trials in which a conditioned stimulus, e.g., a light, and the unconditioned stimulus (acid) are applied in combination. Or if an animal has been given a colored acid, it salivates when shown water of the same color. Also, if a painful stimulus is applied to the animal's paw (unconditioned stimulus) during some form of conditioned stimulation (e.g., sound of a buzzer) the motor reactions (e.g., drawing up the limb, turning the head towards the injured part, etc.) which follow the application of the combined

morphine. In the dog, morphine administration causes vomiting and salivation, followed by sleep. After a series of injections, these effects result from the mere sight of the syringe, or the approach of the attendant who had previously administered the drug.

SECONDARY AND TERTIARY CONDITIONED REFLEXES A second stimulus may be conditioned by linking it up with a conditioned stimulus already firmly established. A defence conditioned reflex, let it be supposed, has been established with an electric shock to the front paw as the unconditioned and a touch upon the hind paw as the

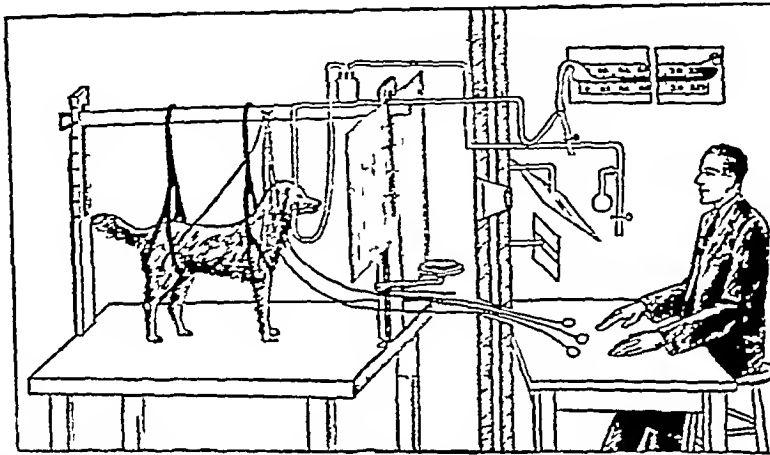


FIG. 691. Illustrating arrangements for experiments upon conditioned reflexes (from Pavlov, *Lectures on Conditioned Reflexes*, International Publishing Company, New York)

stimuli occur, after a few trials, when the conditioned stimulus is employed alone. The painful stimulus may serve also as a conditioned stimulus for a salivary reflex, then, unless the painful stimulus is intense the usual defence reactions are suppressed, salivation alone resulting when the stimulus is applied. The alimentary reflex is in this case stronger than the defence reflex. Pavlov points out that the subordination of the defence reaction by the alimentary reflex is seen when dogs are struggling among themselves for food. Minor injuries (e.g. of the skin) may then be sustained, but they do not evoke a reaction of defence, this is suppressed by the dominant food reflex. On the other hand, if the stimulus (such as one applied to bone) signals a more severe type of injury, or threatens the life of the animal, the defence reaction becomes prepotent.

Of great practical interest is the reflex which becomes established to the repeated injection of

conditioned stimulus. A second neutral stimulus (e.g., the sound of bubbling water) is now applied and withdrawn a few seconds before the application of the primary conditioned stimulus (touching the hind paw) but the unconditioned stimulus (electric shock) is omitted. If the two stimuli are associated in this way a number of times it is found that the second stimulus (sound of bubbling water) has itself acquired conditioned properties, when applied alone the defence reaction occurs. If a third stimulus (e.g., the sound of a tuning fork) is applied a short time before the second, but the primary conditioned and the unconditioned stimuli are omitted, it now, when employed alone, causes the conditioned response. It has not been found possible to establish a conditioned reflex of the fourth order, and conditioned reflexes beyond the second order cannot be established for alimentary conditioned reflexes.

TRACE CONDITIONED REFLEXES In establishing this reflex the unconditioned stimulus is made to follow the conditioned stimulus after an interval, when the reflex has been established the conditioned response follows the conditioned stimulus by an interval of the same duration. For example, a tactile stimulus is applied to the skin for half a minute and then withdrawn, one minute later acid is injected into the mouth. After a number of repetitions it is found that when the tactile stimulus is applied alone, although no secretion occurs during its application, a response follows after a minute's interval. Also, if an animal is fed at regular intervals, say every 30 minutes, it is found that after a series of such feedings secretion occurs spontaneously, thereafter, at intervals of approximately 30 minutes, though no food is given. In these instances the time interval itself has evidently acquired the properties of a conditioned stimulus.

THE BIOLOGICAL SIGNIFICANCE OF CONDITIONED RESPONSES

Conditioned reflexes enter very largely into animal and human behavior. Many such reflexes are developed naturally as the experiences of everyday life become enriched, and associations accumulate. Stimuli arising in the environment are constantly calling forth conditioned responses of various types serving as signals to guide the animal in its choice of action. By training and discipline more complicated reflexes can be established, such as those forming the basis for the tricks of performing animals. In such instances the sound of a word (command) or a movement made by the trainer serves as a conditioned stimulus to some motor reaction of the animal. In the training and education of the child conditioned reflexes also play a prominent rôle. Animals with more highly developed nervous systems are capable of the more complicated reflexes, and though one can scarcely speak of certain conditioned reflexes, such as the secretion of saliva following the flash of a light or the sound of a tuning fork, as an intelligent act, the ability of an animal to develop conditioned responses is, nevertheless, a measure of its intelligence. Conditioned reflexes also play an important rôle in the psychology of sex and, according to Pavlov, in the induction of sleep (p. 1069). Conditioned reflexes which are established under the ordinary circumstances of life are termed *natural*. Those which are established experimentally or by special methods of training are called *artificial*. There is, however, no essential difference between the two. Gantt has enrolled the aid of the conditioned reflex in the investigation of psychi-

atric conditions. In psychoneurotic subjects (organic mental disease) there was impairment or failure to establish conditioned motor responses to an electric shock, whereas in the psychogenic (functional) types, the responses may be impaired but are not abolished. The conditioned reflex is also used to test sensations, especially hearing, in very young children before they can talk.

INHIBITION OF CONDITIONED REFLEXES

Inhibition as applied to conditioned reflexes is divided by Pavlov into *external* or *indirect* and *internal* or *direct*.

A EXTERNAL INHIBITION The conditioned reflex is inhibited by some form of stimulation quite apart from the conditioned stimulus itself. The inhibition arises in a part of the brain other than that in which the conditioned reflex is initiated. For instance, some disturbing factor, a sudden noise, a strange smell, a light, a fresh object in the room, or the entrance of a stranger tends to abolish a conditioned reflex, which in quiet surroundings can be readily elicited. The extraneous stimulus arouses the animal's curiosity and distracts its attention or, in Pavlov's words, evokes an *investigatory reflex*. This purely unconditioned reflex consists of what its name implies—pricking of the ears and turning the eye and head toward the source of the distraction. If the extraneous stimulus is repeated often enough its inhibitory effect is weakened or abolished, the conditioned responses of an animal placed in strange surroundings are at first inhibited but return later. A painful stimulus also sets up an *unconditioned defense reflex*—barking, struggling and other motor reactions which exert an inhibitory effect upon the conditioned response.

B INTERNAL INHIBITION This will be considered under the following headings: (1) *extinction of the conditioned reflex*, (2) *conditioned inhibition*, (3) *inhibition of delay*, and (4) *differential inhibition*.

(1) *Extinction of the conditioned reflex* If a conditioned reflex is repeated a number of times and the unconditioned stimulus (e.g., feeding in the case of an alimentary reflex) always omitted, the response becomes weaker with each repetition, its latent period lengthens progressively and the reflex finally disappears. The reflex is said to have undergone *extinction*. If, however, after every few repetitions of the reflex the conditioned stimulus is followed by the unconditioned stimulus, extinc-

tion does not occur. The former stimulus is then said to have been *reinforced* by the latter. For example, if a conditioned alimentary reflex has been established to the sound of a buzzer, the application of the latter alone calls forth at first a prompt and ample secretion of saliva. After a variable number of repetitions in which the conditioned stimulus is not followed by feeding, i.e., not reinforced, the secretion becomes less each time and finally ceases. After an hour or two the reflex recovers spontaneously. Extinction is due to inhibition of the cortical elements of the reflex. That is not due simply to fatigue of the salivary gland or of the medullary centers is shown by the fact that after complete extinction, reinforcement of the conditioned stimulus causes the reestablishment of the reflex. Moreover, the reflex continues at full strength after a great number of repetitions if it is followed every few times by reinforcement.

After a reflex has undergone extinction, some external stimulus may temporarily remove the inhibition. That is, just as an extraneous stimulus can inhibit the excitatory phase of a conditioned reflex, so also can it inhibit the inhibitory state. This is called *disinhibition* or the "*inhibition of inhibition*"

(2) *Conditioned inhibition*. If, after a positive conditioned reflex has been firmly established another stimulus is combined with the conditioned stimulus for a number of trials but reinforcement is always omitted, then, though the conditioned stimulus still causes the customary response when applied alone (and regularly reinforced), it is quite ineffective when in combination with the extra stimulus. For example, an alimentary conditioned stimulus is established to the beat of a metronome. Later, a buzzer is sounded with the metronome, but the combination is not followed by feeding. After a series of such trials the metronome causes a secretion, but the combined stimuli, metronome plus buzzer, are without effect. The sound of the buzzer is termed a *conditioned inhibitor*, and the inhibitory effect which it produces, a *negative* or *inhibitory conditioned reflex*. Usually, in order to demonstrate the inhibitory effect, the primary conditioned stimulus and the conditioned inhibitor must overlap.

(3) *Inhibition of delay*. If during the establishment of a conditioned alimentary reflex, the conditioned stimulus is continued for only a brief period, 1 to 5 seconds, before the conditioned stimulus is applied, then the conditioned response (secretion of saliva) follows almost immediately

upon the commencement of the conditioned stimulus. That is, the reflex has a very short latent period. If after such a *simultaneous reflex* has been established, it is repeated day after day, but the conditioned stimulus is continued a little longer each time before the reflex is reinforced, the latent period becomes lengthened in proportion to the interval between the application of the two stimuli. The almost simultaneous reflex has been converted into a *delayed reflex*. In other words, postponement of reinforcement has caused the conditioned response to be inhibited during the first part of the action of the conditioned stimulus, during the latter part of the action of the conditioned stimulus the secretion of saliva commences and increases in amount up to the moment when reinforcement ordinarily would have occurred.

(4) *Differential inhibition*. This will be described in the next section.

The examples of internal inhibition just given show the high degree of discriminative and adaptive powers of which the cerebral cortex is capable. Though such adjustments are purely automatic they are effected with great delicacy and an apparent purpose. In the case of extinction, for instance, the futility of secreting saliva for food which does not follow appears to have been "*recognized*". The purpose in conditioned inhibition is quite as evident as in positive conditioned reflexes and the inhibition of delay is clearly an adjustment which economically times the secretory response to the moment when food is "*expected*".

It is not in these special instances that internal inhibition occurs, for all positive conditioned reflexes, though reinforced regularly, undergo inhibition if repeated over a period varying in different animals from weeks to months or even years. They become progressively weaker, the latent period lengthens out and they ultimately disappear. The tendency of conditioned reflexes to undergo inhibition is an inherent property.

ANALYZING AND SYNTHESIZING FUNCTIONS OF THE CEREBRAL CORTEX

Of the numberless agencies in the environment to which the organism is exposed the great majority might be termed neutral in that they exert neither a beneficial nor an injurious effect. The actions of others are either of definite physiological value or detrimental to the animal's existence. Through the analyzing mechanism possessed by the nervous system the stimuli to which the latter types of agent give rise are given conditioned

properties. Such stimuli are picked out to serve as signals for reactions on the part of the animal appropriate to the respective agents (beneficial or noxious), they are therefore of the utmost biological importance. The cerebral cortex also possesses synthesizing mechanisms whereby individual stimuli are fused into conditioned complexes (p. 1048).

THE ANALYZERS

Pavlov divides the neural mechanism of the organism upon which the discriminative faculties depend into a number of *nervous analyzers*. These are constituted of the nerves of special sense, and the afferent nerves of the joints and skeletal muscles, together with their respective receptors and central connections. Thus he speaks of *visual*, *auditory*, *olfactory*, *gustatory*, *cutaneous* and *motor analyzers*. The receptor of each class is especially responsive to its own type of stimulus—light, sound, etc. The central part of each analyzer, i.e., its terminations in the cerebral cortex, is capable of a very fine discrimination between the different intensities and qualities of stimuli within its own class. The visual analyzer, for example, discriminates between the intensity and quality of different visual stimuli, the auditory analyzer between the intensity, pitch and quality of sounds, the motor analyzer between the various messages (proprioceptive) received from the muscles and joints—and so on for the other analyzers.

In the past, the study of these analyzers, i.e., of the sense organs, has been based very largely upon subjective data gleaned from experiments upon the human subject. The discovery of conditioned reflexes, however, has provided a reliable method for the study of the analyzing functions which being purely objective can in consequence be employed in animal experimentation, the conditioned salivary secretion, for example, is a reaction which readily lends itself to precise measurement and timing.

GENERALIZATION AND DIFFERENTIATION

If a conditioned reflex is established, say to the sound of a tuning fork of 800 cycles per second (c.p.s.), it is found that tones somewhat higher or lower in the scale have also acquired conditioned properties. Also, after a conditioned reflex has been established to a tactile stimulus applied to a certain definite skin area, the stimulation of neighboring areas is also effective. The response,

however, becomes weaker the farther away from the original area that the stimulus is applied. This *generalization of stimuli*, as Pavlov calls the phenomenon, is seen also in the case of olfactory, visual and other analyzers. If, however, the original definite stimulus, for example the tone of 800 c.p.s., is always followed by reinforcement while other tones having a higher or lower frequency are employed without reinforcement, then only the tone of 800 c.p.s. evokes a response. The allied stimuli are said to have undergone *differentiation* from the primary stimulus. Pavlov ascribes the phenomenon to a form of internal inhibition—*differential inhibition*. He believes that originally the excitatory process in the cortical part of the analyzer is widespread, but through the antagonism offered by the internal inhibition set up by non-reinforcement of the allied stimuli, it becomes localized to only a minute cortical area corresponding to the receptors affected by the primary stimulus.

The degree to which differentiation between various types of stimulus can be developed is a measure of the analyzing ability of the cerebral hemispheres and is often amazing. The following examples are taken from Pavlov's monograph.

1. **AUDITORY STIMULI** (a) *Differentiation of pitch*. Primary stimulus 800 c.p.s. Differentiated stimulus 812 c.p.s. (b) *Differentiation of rhythm*. Primary stimulus 120 beats per minute of a metronome. Differentiated stimulus 118 beats per minute of a metronome (Andreyev). (c) *Differentiation of intensity*. The difference in the intensity of two sounds was so slight that it was detectable by the human ear only when one stimulus was followed immediately by the other. Differentiation was perfectly effected by the dog when the two stimuli were separated by an interval of 17 hours.

2. **VISUAL STIMULI** (a) *Differentiation of direction of movement or of the position of an object*. Primary stimulus, clockwise rotation of a disc. Differentiated stimulus, anticlockwise rotation of the disc. (b) *Differentiation of figures and shapes*. Some of the figures which were differentiated are shown in figure 69.2. A luminous circle thrown upon a screen was readily differentiated from a series of ellipses of the same luminosity, the series started with one having its axes in the ratio of 1:2, of the remainder each successive one approached a little nearer to the circular shape. Differentiation just failed when the ratio of the axes was 8:9. (c) *Differentiation of luminosity*. Two shades of gray which to the human eye appeared exactly the same, even when viewed simultaneously, were perfectly differentiated by the dog when an interval of a minute separated the primary from the differentiated stimulus. (d) *Differentiation of*

colors failed in all but one animal investigated, and even it gave a doubtful result. Color vision in the dog is therefore either absent or very rudimentary.

3 DIFFERENTIATION OF CUTANEOUS AND PROPRIOCEPTIVE STIMULI. Differentiation was obtained for various types of tactile stimuli, e.g., contact with rough or smooth surfaces, pressure with blunt points arranged in different patterns, scratching with a small brush in different directions. Differentiation was also demonstrated between stimuli applied to different areas, for variations in temperature, and between various passive movements, e.g., flexion of ankle as against extension.

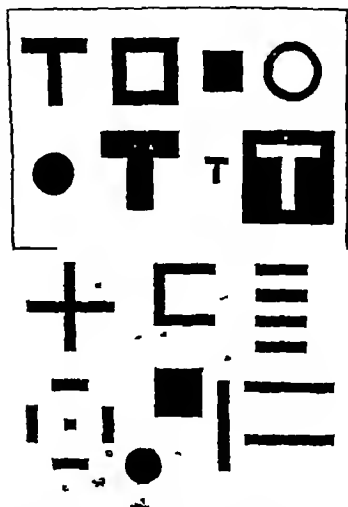


FIG. 69.2 Examples of different figures which were successfully differentiated in experiments upon a dog. The letter T, shown in the upper left hand corner of the figure, served for the positive stimulus, the other black figures and the white letter T were differentiated from the positive stimulus. In another dog the white cross was the positive stimulus from which the other white figures were differentiated (from Pavlov, *Conditioned Reflexes*, Oxford University Press)

4 DIFFERENTIATION OF OLFACTORY AND GUSTATORY STIMULI. Corresponding differentiations were established for various odors (e.g., vanillin, amyl acetate, camphor, etc.) and for taste sensations e.g., meat powders, sugar, cheese, etc.

Conditioned vasomotor responses have been established in human subjects, the unconditioned response (vasoconstriction of the opposite hand) being induced by immersion of one hand in ice-cold water. Ringing of a bell, a light pattern or a word spoken aloud by the experimenter and whispered by the subject were used as conditioned stimuli.

"Experimental neurosis" When an animal is presented with a problem which requires a fine degree of differentiating ability, i.e., when a con-

flict between inhibitory and excitatory processes is set up, either inhibition or excitation may gain the upper hand with the suppression of the opposite process. In the former event the animal may become drowsy and fall into a deep sleep (p. 1069), whereas in the latter a nervous disturbance may develop during which even gross differentiation cannot be accomplished. To give an example, a dog was required to discriminate between two visual stimuli, a circle and an oval, the former being followed by feeding (positive stimulus), while the latter was not reinforced (negative stimulus). Differentiation was made more and more difficult by making the oval at each successive trial more nearly circular. The experiment proceeded smoothly until the axes of the two figures had a ratio of 8:9. The animal then became fractious, howled and whined in its cage, struggled with its harness and became incapable of discriminating between any oval shape and a circle. Not until after a prolonged rest was the animal's power of discrimination between the two shapes restored, but upon being then confronted with the same problem the nervous disorder returned "*Experimental neurosis*", as this state was termed by Pavlov, has also been induced in sheep and pigs by Anderson and Liddell. In some instances the condition was permanent, and Gantt has reported such a state which has persisted for 12 years. The neurosis in this latter instance developed as a result of a conflict aroused in the differentiation of two auditory conditioned stimuli—two closely similar tones. These observations upon experimental neuroses have important psychiatric implications and may give a lead in the interpretation of certain nervous conditions in the human subject. Sexual excitement, as manifested by erections, are not uncommonly associated with the neurosis (see Gantt).

SYNTHESIS OF STIMULI

The development of the conditioned response is, in itself, evidence of the synthesizing or associative ability of the cortex. Further evidence is afforded by experiments with compound stimuli, i.e., the fusion of separate individual stimuli, whether these act upon the same or upon different analyzers, into a conditioned complex.

As an example of simultaneous stimuli acting upon the same analyzer,—an alimentary conditioned reflex was established to a chord of three tones of equal intensity but having a frequency of 85, 256 and 786 c.p.s., respectively. Later each

tone when sounded separately caused a response. The responses to the different tones were approximately equal but weaker than that caused by the chord.

A conditioned reflex may be established to a compound stimulus made up of two stimuli acting *simultaneously* but upon different analyzers. If the component stimuli are then applied separately and without reinforcement, it is found that one of the pair is effective but not the other. For example, a conditioned reflex to acid was established to a compound stimulus consisting of the simultaneous application of a tactile stimulus and a thermal stimulus of 0°C. The tactile stimulus when applied alone was found to be about as effective as the compound stimulus, whereas the thermal stimulus was quite ineffective. The results of experiments in which the component stimuli of the compound stimulus were of unequal strengths, and applied to the same analyzer, indicate that the factor determining the effectiveness of the stimuli, when applied separately, is their relative intensity. The thermal stimulus used in the foregoing experiment may therefore be regarded as being the weaker of the two stimuli. Visual stimuli are also weaker than those acting upon the auditory analyzer. For example, after a conditioned reflex had been established to a compound stimulus made up of a tone and a light, the tone by itself caused a response, whereas the visual stimulus was quite ineffective. However, the weaker stimulus undoubtedly plays its part in the combination, for if the stronger component is applied repeatedly without reinforcement by the unconditioned stimulus, but the compound stimulus is constantly reinforced, the stronger stimulus by itself becomes ineffective, whereas the compound stimulus retains its full effect.

The cortex is also able to synthesize *successive* stimuli into a compound conditioned stimulus. For instance, a flash of light (L), a cutaneous stimulus (C) and the sound of bubbling water (S) when applied in the order (L—C—S) were compounded into a positive conditioned stimulus. The reverse order (S—C—L) after differentiation (by non-reinforcement) was entirely without effect.

Similarly, four tones having vibration frequencies of 290, 325, 370 and 413 c.p.s. respectively when sounded in this order (1, 2, 3, 4) were successfully differentiated by a dog from all other sequences (e.g., 4, 3, 2, 1, 4, 3, 1, 2, 4, 1, 3, 2, 4, 2, 3, 1, etc., etc.).

The series of sounds in a word (e.g., a dog's

name) or in a command is a familiar example of a successive compound conditioned stimulus.

IRRADIATION AND CONCENTRATION

These features are best illustrated by citing an actual experiment. The skin of the hind limb was stimulated at 5 separate places. One of these situated upon the paw was differentiated by non-reinforcement in the usual way from the other four. That is, the place upon the paw was given inhibitory properties (p. 1060). The other four places, which through reinforcement were given positive conditioned properties, were placed at progressively further distances along the limb from the inhibitory place. It was found that if the latter was stimulated three times in succession, and one minute after the last stimulation, the positive place nearest the inhibitory place was then stimulated, there was complete failure of the usual conditioned response. Stimulation of the next (second) excitatory place one minute after the last application of the inhibitory stimulus was followed by a response half as great as usual. The remaining two places (third and fourth) gave a normal or somewhat greater than normal response (see positive induction, p. 1064). When the number of the previous stimulations of the inhibitory place was increased, and the interval following the last one shortened to $\frac{1}{2}$ minute, all four excitatory places gave a reduced response. As time elapsed the furthest place was first freed from the inhibitory influence, and the others in the order of their positions from the inhibitory place. The point nearest to the latter in some cases did not give a full response until the lapse of 10 minutes after the application of the last stimulus to the inhibitory place. It can also be shown that if, by means of extinction, inhibitory properties are given to one of the four excitatory places the inhibitory process spreads to involve the others.

These results are taken to indicate that each stimulated area has its circumscribed representation in the cortex. Inhibition initiated in the cortical cells to which the inhibitory place is projected *irradiates* to involve the cortical projection areas of the positive places. The gradual release of the cortical areas from the inhibitory influence, and its retreat to the original cortical cells corresponding to the inhibitory place is called the *concentration of inhibition*. Irradiation and concentration are antagonistic processes. These processes can also be shown for the acoustic analyzer and they most probably exist for other analyzers.

as well. The following is an illustration of irradiation within the acoustic analyzer. Positive conditioned reflexes were established respectively to the beat of a metronome, to a musical tone and to a buzzing sound. When the reflex to the buzzing sound or to the metronome was extinguished the inhibitory process spread to involve the reflex to the musical tone.

Irradiation is not always confined to the analyzer in which the inhibitory process originates. It can be shown that the inhibitory process initiated in one analyzer spreads to others and may involve the entire cortex. For example, a positive alimentary conditioned reflex was established to a tone of 4000 c.p.s., a note a semitone lower than this was given inhibitory properties. When the inhibitory stimulus was applied a short time before a visual conditioned stimulus, the reflex which had been established to the latter, as well as that established for the positive auditory stimulus (tone of 4000 c.p.s.) was inhibited.

Irradiation and concentration of excitation occur in a manner analogous to that described for inhibition. Generalization (p. 1061) is ascribed to the irradiation of excitation from the primary cortical focus. Differential inhibition, on the other hand, antagonizes the excitatory process and concentrates it within the original cortical area.

INDUCTION

Induction is a feature of conditioned reflex action analogous to the phenomenon of the same name described as occurring in spinal reflexes (p. 950), namely, the increase of inhibition caused by a preceding state of excitation (*negative induction*) or an increase of excitation caused by a previous state of inhibition (*positive induction*). Induction is therefore a reciprocal process.

Negative induction: The inhibitory properties of a stimulus developed by means of differentiation can be readily destroyed again (disinhibition) by repeated reinforcement. If, however, each application of the inhibitory stimulus is preceded by a positive one and both are reinforced the inhibitory properties of the former are strengthened and can be abolished only after a great number of trials. Thus, a metronome beating at a rate of 120 beats per minute served as a positive conditioned stimulus and a rate of 60 beats per minute was differentiated (differential inhibition, p. 1060). The inhibition was then almost abolished (disinhibition) by reinforcement but was restored again when the differentiated

stimulus (rate of 60 beats per minute) was preceded by the positive stimulus (rate of 120 beats per minute). The data are shown in table 96, given by Pavlov.

TABLE 96

TIME	CONDITIONED STIMULUS APPLIED DURING 30 SECONDS	SALIVARY SECRETION IN DOGS DURING 30 SECONDS DROPS
11 25 a.m.	60 beats	0
11 30 a.m.	60 beats	0
11 42 a.m.	60 beats	3
11 49 a.m.	60 beats	4
11 56 a.m.	120 beats	8½
12 06 p.m.	60 beats	0

All the stimuli were accompanied by reinforcement

Positive induction in its simplest form may be illustrated as follows. A positive conditioned reflex was established to stimulation of the forepaw. A stimulus to the hind paw was given inhibitory properties by differentiation. When the excitatory place (forepaw) was stimulated immediately after the application of the stimulus to the inhibitory place, the positive conditioned reflex was enhanced (salivary secretion increased by 50 per cent). The excitability of the cortical area receiving impulses from the forepaw had evidently been increased as a result of inhibition of the cortical area corresponding to the hind paw. Thus the excitability of the cortex in proximity to the inhibited area is enhanced for the moment but later becomes depressed as a result of the irradiation of the inhibitory process. When dealing with irradiation it was pointed out that the inhibitory effect was demonstrable for several seconds or minutes after the inhibitory place had been stimulated. Positive induction occurs during the application of the inhibitory stimulus. Pavlov holds the view that these two processes, positive induction and irradiation of inhibition (or negative induction and the irradiation of excitation), are constantly interacting with one another and spread wave-like over the cortex. Positive or negative induction precedes, respectively, the spread of the inhibitory or excitatory process.

THE EMPLOYMENT OF CONDITIONED REFLEXES IN THE STUDY OF CORTICAL FUNCTION Information concerning the localization of function in the cortex was obtained by, first, firmly establishing a conditioned reflex involving one or other analyzer, then excising a certain portion of the cortex

and studying the effects of the operation upon the conditioned response

REMOVAL OF THE ENTIRE CORTEX

In the dog removal of the entire cortex results in the complete and permanent loss of all conditioned responses, natural or artificial, new conditioned reflexes cannot be established. The decorticated animal responds to crude *unconditioned* stimuli, e g, a bright light or sound, its reactions being frequently those of resentment or anger (p 1027). The unconditioned salivary reflex is lost at first but eventually becomes stronger than the normal.

REMOVAL OF A LIMITED PORTION OF THE CORTEX

Removal of a limited portion of the cortex, e g, the temporal or occipital lobes, causes the disappearance for a time of all "artificial" conditioned reflexes, and sometimes of the "natural" ones as well. Unconditioned reflexes, which are dependent upon subcortical centers, remain in abeyance for only a few hours or may not disappear at all. Complete recovery of conditioned reflexes dependent upon analyzers, other than the one which has been directly injured, occurs after a few days, analyzers nearest to the excised area recovering later than those farther removed. Incomplete recovery of the injured analyzer eventually takes place.

THE ACOUSTIC ANALYZER is centered in the temporal lobe. Extirpation of both temporal lobes is followed by a very temporary loss, if any, of the auditory unconditioned reflexes (e g, pricking of the ears to sound). The auditory conditioned reflexes may not return for several weeks or even months, and are never fully restored. The finer and more discriminating reactions are permanently lost. Discrimination (p 1061) of the intensity, pitch and quality of sound is possible, but the powers for more complicated analyses and syntheses are not regained. The animal, for example, never answers to its name, and the differentiation of other types of compound auditory stimuli (p 1063), e g, a descending from an ascending scale of tones, cannot be established.²

² An experiment upon the peripheral portion of the acoustic analyzer may be mentioned here, since it has a bearing upon Helmholtz' theory of hearing. Destruction of the organ of Corti in the upper part of the cochlea (receptors for lower tones) resulted in a loss of conditioned reflexes which had been established to tones of the lower part of the musical scale (frequencies lower than 600 d v per minute). Reflexes to higher tones were unaffected.

THE VISUAL ANALYZER is situated mainly but not exclusively in the occipital lobe. According to Pavlov the visual analyzer is probably spread over the entire cortex. Complete bilateral extirpation of the occipital lobes causes effects upon the visual conditioned reflexes corresponding to those upon the auditory reflexes which follow removal of the temporal lobes. Conditioned responses to changes in the intensity of illumination (e g, switching on a high power lamp) established before operation returned upon the fifth day following the extirpation of the occipital lobes as well as of a considerable area of the cortex lying anterior to them. Differentiation was established between a luminous cross and a circle illuminated to the same intensity. It was found impossible, however, to establish conditioned reflexes to specific objects, for the reason, most probably, that the appearance of an object changes with the illumination, with its distance from the eye and the angle from which it is observed. Its recognition under different conditions of lighting, position, etc, therefore requires a much greater power of analysis and synthesis than that required for the differentiation of a flat luminous shape. The animal deprived of its occipital lobes is able, nevertheless, to avoid objects placed in its path. They are detected simply by the changes in illumination (lights and shadows) which their presence creates.

The findings described above, namely, that complete decortication in the dog abolishes all conditioned reflexes, and that removal of the temporal or occipital lobes impairs, respectively, auditory, and visual conditioned reflexes, but does not destroy them permanently and completely, lead to the following conclusions concerning the functions of the cortex, so far at any rate as the auditory and visual senses are concerned. (a) Though rudimentary reactions can be carried out through subcortical connections, the processes underlying conditioned reflexes are dependent upon the activity of the cerebral cortex. (b) The highest powers of analysis and synthesis are localized to definite areas, each of which may be regarded as the "nucleus" of a given analyzer. (c) The analyzer, however, is not rigidly confined to this area but extends into other cortical regions in which less complicated types of analysis can be undertaken. Experiments involving other analyzers have yielded evidence pointing in the same direction.

These results of cortical extirpation cannot be directly applied to the human subject since the

functions of subcortical centers of hearing, sight, etc., in lower animals have not, as in man, been so largely usurped by the cortex. Nevertheless, Pavlov's observations are, broadly speaking, in harmony with modern conceptions of cortical functions (see p 1038). A given sensory function, though primarily "centered" in a certain cortical area, is, through association tracts, dependent for its full development upon other areas as well.

THE EFFECTS OF DRUGS UPON CONDITIONED REFLEXES

Caffeine and strychnine increase the effects of positive conditioned stimuli and weaken internal inhibition. After a dose of less than 1 grain (0.025 to 0.05 gram) of the former drug, extinction (inhibition) of a conditioned reflex is effected with the greatest difficulty. *Bromides* act by strengthening internal inhibition, and not by directly depressing excitatory processes, under their influence the extinction of positive conditioned reflexes is facilitated. *Alcohol* in moderate doses weakens internal inhibition. Andreyev and Fugsley have found that the *hypercalcemia* resulting from parathormone or ergosterol overdosage causes an exaggeration of inhibitory processes, the after effect of inhibitory stimuli is enhanced and extinction of positive conditioned reflexes accelerated.

THE PHYSIOLOGY OF SLEEP

THE DEPTH OF SLEEP

The depth of sleep is not constant throughout the sleeping period but varies from hour to hour. Experiments upon man in which auditory stimuli were employed to arouse the subject at different times, or the movements of the sleeper were recorded (the depth of sleep being assumed to be inversely related to the amount of muscular movement) indicate that the depth of sleep follows a characteristic curve. In most adults sleep deepens rapidly to the end of the first hour, after which it lessens sharply for a time, and then more slowly till the time of waking. In children the sleep curve shows two maxima, i.e., two periods of the deepest sleep, one of these is reached in the first or second hour, the other between the eighth and ninth hours, the curve then falls rapidly to the time of waking. Generally speaking, sleep taken during the daytime is lighter than that during the night. Deep sleep is dreamless, dreams occur only during light sleep and chiefly in the period which just precedes waking. In sleep, unconsciousness is not uniform for all senses, the depth of sleep is greatest for the sensations of smell and least for those of pain, hearing and touch.

The *sleep requirement* of different persons varies widely, it also alters with age. The following are average figures for the hours of sleep required at different periods of life.

	hours
New-born infant	18 to 20
Growing children	12 to 14
Adults	7 to 9
Old persons	5 to 7

PHYSIOLOGICAL CHANGES ACCOMPANYING SLEEP

During sleep most bodily functions are reduced to their basal levels. The *blood pressure* is lowered, the systolic pressure showing a decline of from 10 to 30 mm Hg. The lowest level is reached about the fourth hour of sleep, and remains at this level until a short time before awakening, when the pressure commences to rise again. MacWilliam found that if the sleep was disturbed by exciting dreams the blood pressure might be elevated well above the normal waking level, of 125 or 130 mm Hg. The *pulse rate* is slowed by from 10 to 30 beats. The *metabolic rate* is reduced by from 10 to 15 per cent below the basal level and the *rectal temperature* by a fraction of a degree Fahrenheit. The *heat-regulating mechanisms* are depressed. The *respirations* are slowed as a rule, and are said to become more costal in character, they also tend to become irregular or periodic. *Muscle tone* is minimal, the knee jerk is abolished, and a positive Babinski may be present. The thresholds for most *somatic reflexes* are definitely raised. Vasomotor reflexes, however, are more active. In most animals the *righting reflexes* are abolished. The pupils are usually constricted, the *light reflex* is retained. The eyeballs are turned upwards and outwards. *Urine volume* is reduced, but the absolute excretion of urinary phosphate is increased, and the specific gravity raised. The *secretion of the sweat glands* is considerably increased, according to Hartnidge the quantity of fluid lost per hour in sleep is nearly equal to that lost during a corresponding period of strenuous muscular exercise. *Plasma volume* is reduced by about 10 per cent. *Gastric secretion* is increased or little altered during sleep. The contractions of the empty stomach continue and may be more vigorous than usual, the rate of digestion is about the same as during the waking state. *Lacrymal* and *salivary* secretions are reduced. The EEG during sleep is described in chapter 68.

It may be of interest to give an account of the effects upon the nervous system of prolonged

wakefulness In Kleitman's human experiments the subjects were kept awake for periods ranging from 60 to 114 hours The knee jerk remained unaffected but disappeared promptly when the subject went to sleep at the termination of the period of forced insomnia, and a positive Babinski was obtainable The latter was attributed to the establishment of a block in the corticospinal pathway The *pupillary* response remained brisk throughout the wakeful period There seemed to be little impairment of the mental processes, and the reactions to auditory and visual stimuli were as prompt as usual The threshold for pain stimuli was definitely lowered, whereas that for touch was unaltered The power to maintain *equilibrium*, as judged by the ability to stand with the eyes closed, was grossly impaired This defect was attributed to neuromuscular fatigue and the consequent reduction in muscle tone, rather than to any impairment of labyrinthine function itself

Tyler, in more recent experiments upon some 600 human subjects who went for periods of up to 112 hours without sleep, found no significant changes in blood chemistry, hemoglobin, red or white cell count, body weight or temperature, and only slight changes in blood pressure, respirations and heart rate Notable psychological changes, however, were observed, e.g., loss of memory, irritability, inattention and hallucinations or illusions These alterations in behavior were noticeable after from 30 to 60 hours of sleeplessness, and, though mild in most subjects, in a few, they were severe and resembled those of acute schizophrenia

The period of wakefulness which would be lethal for the human subject is not definitely known Dogs may die after being kept awake continuously for 14 days, though they may survive for much longer periods Young animals are much more susceptible to loss of sleep than older ones Changes in the nerve cells of the cortex, e.g., chromatolysis and shrinkage of the cell bodies, have been described as resulting from prolonged periods of enforced wakefulness in animals

THEORIES OF SLEEP

The cause and nature of sleep have aroused speculation from the time of the Greek philosophers, but though hypotheses are many and often ill-founded, facts which might throw light upon the underlying processes are few and difficult to obtain A discussion of only some of the many theories will be undertaken

(1) **NEURON THEORY** (Lépine, Duval) This theory arose from the demonstration by Cajal that there was not anatomical continuity between adjacent neurons but merely points of contact, for which the term *synapse* was later suggested by Foster The neuron theory postulated that the function of the cells of the higher cortical centers was suspended as a result of the retraction of the dendritic processes, and the consequent break in contact between neurons Though some histological evidence was cited in support of this conception, the theory was mainly speculative

(2) **CEREBRAL ISCHEMIA THEORY** Howell suggested that fatigue of the vasomotor center with consequent vasodilatation of the peripheral vessels, especially of the skin, and reduction in cerebral blood flow was the primary change responsible for the onset of sleep The flushed skin of the sleeping subject, the fall in blood pressure and the well-known feeling of drowsiness following a meal (which presumably was the result of the diversion of blood to the splanchnic area) lent plausibility to the theory From ancient times the carotid artery (*karoō* = I sleep) has been believed to be connected in some way with the mechanism of sleep, for it was recognized that compression of this vessel was not uncommonly followed by unconsciousness In later times the loss of consciousness following pressure upon the neck was attributed to vagal stimulation, with consequent inhibition of the heart and a reduction in intracranial blood flow We know now that the unconsciousness is due to the fall in blood pressure brought about through the carotid sinus mechanism

The unconscious states caused by a reduction in cerebral blood flow are not, however, akin to normal sleep Vulpien observed some years ago that, though stimulation of the cervical sympathetic in animals caused cerebral ischemia, sleep was not induced Moreover, Gibbs has shown by means of an electrically heated stylet (p 341) that no diminution in blood flow through the brain occurs during natural sleep It is possible, nevertheless, that there may occur a reduction in the blood supply to a limited area of the brain whose activity is essential for the waking state A limited vascular change of this nature might not be revealed by observations upon the total intracranial blood flow With regard to the carotid sinus and its possible relationship to the sleep mechanism, a recent interesting observation should be mentioned Weiss and his associates describe states resembling normal sleep as resulting from pressure

upon the sinus in certain susceptible subjects. The notable feature of this reaction is that unconsciousness is not accompanied by a significant change in heart rate, blood pressure or in the blood flow through the brain. In some instances there was actually a rise in blood pressure preceding and during the unconscious state. The effect of carotid sinus stimulation in these instances appears to have a purely nervous basis, the sensory part of the mechanism being the sinus nerve and the receptors in the sinus wall. The brain center or centers involved are unknown. Koch also reported in 1932 that in dogs recovering from anesthesia, stimulation of the carotid sinus caused the immediate cessation of muscular movement, the head and tail of the animal gradually hung more and more limply, or as Koch expresses it, "The animal sinks loosely together, often on its side, to be as in sleep."

(3) **CHEMICAL THEORIES** Several chemical theories of sleep have been proposed. One of the earliest of these was that fatigue products, especially lactic acid, formed in the tissues generally, acted by depressing the function of the cortex. Against this view is the well-known fact that one need not be fatigued in order to sleep, on the other hand, a person may be unable to sleep though utterly fatigued. Furthermore, the brain tissue actually derives energy from the oxidation of lactic acid.

A more recent chemical theory is that of Pieron. This observer claims that a substance, which he terms *hypnotoxin*, is produced by the brain tissue and acts as a soporific. He claims that the cerebrospinal fluid of a dog killed during sleep induces sleep when injected into another animal. Kroll makes similar claims, stating that an acetone extract of the brain of a sleeping or hibernating animal will cause sleep when injected into another. Holmes has been unable to confirm Kroll's findings. An extract prepared by Kroll's method either from sleeping or waking animals was found to be lethal through its toxic effect upon the heart. Ivy and Schnedorf have repeated Pieron's experiments and confirmed his observation. They found that the injection of cerebrospinal fluid of a dog kept awake for several days induced a state of depression resembling deep sleep in rested dogs, when introduced into the cisterna magna or cerebral ventricle. The effect, however, is probably due, not to the presence of a sleep-inducing substance in the "fatigued" cerebrospinal fluid, but to a rise in intracranial pressure, since the injection of cerebrospinal fluid from rested dogs had a depressing action not considerably less than that of the "fatigued" fluid. No evidence in support of the theory (Dikshut) that *acetylcholine* liberation by the brain tissue is a factor in sleep was secured by these observers.

Zondek and Bier have advanced an interesting, but unsubstantiated theory in which the pituitary plays a leading rôle. They state that the pituitary during the waking state has a higher concentration in bromine than any other tissue but that during sleep the bromine concentration of the gland diminishes while that of the medulla increases. These observers therefore believe that sleep is induced through the liberation of a bromine compound from the hypophysis. They have named this substance *bromhormone*. The evidence upon which this theory is based is far from convincing.

Other chemical theories, based upon the supposed affinity of a subcortical area for certain ions, especially calcium whose depressant action upon nervous tissue is well known, have been advanced, but little substantial evidence has been secured for their support.

(4) **THE DIENCEPHALON AND SLEEP** Several observations both clinical and experimental point to the existence of a sleep center in the diencephalon. Hypersomnolence is a frequent accompaniment of tumors of the structures in the floor and walls of the third ventricle, or of inflammatory lesions involving the hypothalamic region. Hess claims to be able to cause sleep in animals by mild electrical stimulation of the diencephalon towards the anterior end of the cerebral aqueduct, and Gagel has induced sleep in human subjects during operations by mechanical stimulation of the posterior part of the hypothalamus. Hess also reported that ergotamine injected directly into the third ventricle induces sleep. Ergotamine paralyzes the motor and secretory fibers of the sympathetic. Since the parasympathetic and sympathetic centers are apparently situated in the hypothalamus, Hess argued that the drug by suspending the activity of the sympathetic center caused a preponderance of parasympathetic effects. Sleep, he concluded, was a parasympathetic function. He also draws support for his theory from certain manifestations of parasympathetic activity, namely, the pupillary constriction, bradycardia and vasodilatation which accompany sleep. Though this observer's idea that sleep is a parasympathetic function requires further experimental support before it can be accepted, the evidence for the participation of the hypothalamus in the sleep mechanism is very strong. Yet, contrary to the idea that sleep is caused by excitation of some part of the hypothalamus, other investigators believe that it results from the *depression* of hypothalamic activity. The hypothalamus is thus regarded as containing a waking center, sleep following its destruction or

inhibition³ Ranson and his associates, for example, have reported that sleep can be readily induced in cats by lesions placed in the posterior and lateral part of the hypothalamus or by injury confined to the mammillary bodies, but that it did not result from the stimulation of any part of the hypothalamus. Harrison also found that electrical stimulation of the hypothalamus caused somnolence only when the current exerted a destructive action.

Whether the thalamus plays a rôle in the sleep mechanism is unknown. Though unconsciousness may result from manipulations in the region of the thalamus, and a stuporous state may be associated with a thalamic lesion, true sleep has not been produced by either stimulation or injury.

(5) **PAVLOV'S THEORY** Pavlov believes sleep and internal inhibition to be essentially one and the same process, i.e., sleep is simply the spread (irradiation) of internal inhibition over the entire cortex with the subsequent involvement of subcortical levels, and internal inhibition confined within the boundaries of a single analyzer is a localized sleep. He was led to this conclusion by the behavior of animals during his investigations of conditioned reflex action. Drowsiness and sleep were frequent accompaniments of all forms of internal inhibition, e.g., inhibition of extinction and of delay (p. 1060), conditioned and differential inhibition, or the inhibition which ensued spontaneously after the repetition of positive conditioned reflexes over a long period of time. An animal, for example, which is quite alert during the establishment of a reflex to a definite musical tone becomes drowsy and falls asleep in the stand during attempts to develop differentiation (p. 1061) of a closely similar tone, its muscles relax, it may snore loudly and other positive conditioned stimuli fail to awake it. Moreover, drugs such as caffeine (p. 1066) which reduce internal inhibition, and those such as bromides which increase it, have corresponding effects upon the mechanism of sleep. As mentioned previously, positive conditioned reflexes after having been repeated over a long period ultimately undergo inhibition. At this stage in the investigations the experiment is frequently terminated by the animal falling

asleep. Experiments involving the use of thermal and tactile stimuli are most frequently interrupted by the onset of sleep, auditory conditioned stimuli are the least likely to have this effect.

Protracted, mild stimulation of an extraneous nature was also found to cause cortical inhibition and lead to sleep. It has been mentioned (p. 1059) that an extraneous stimulus induces inhibition, through setting up an investigatory reflex, upon repetition, the inhibitory effect disappears, and the conditioned responses return. With further repetition, however, the extraneous stimulus again causes inhibition, this time, it exerts, of itself, a direct inhibitory effect upon the cortex.

Pavlov's theory has much to recommend it. The drowsiness which results from some oft-repeated form of monotonous stimulation, e.g., reading or being read to in a low even voice, a dull lecture, or boredom from whatever cause, is well known. Also, the preparations for sleep—the various agencies in a familiar environment—probably serve themselves as inhibitory conditioned stimuli. A dog, for example, which has fallen asleep during previous experiments may do the same when merely brought into the room where the experiments have been performed or when preparations are being made to repeat them. The customary hour for retiring probably acts also as a time conditioned stimulus.

For a critique of Pavlov's theory see Denny-Brown.

In some experiments described by Pavlov the inhibitory process involved the cortex but did not descend to subcortical levels governing equilibrium and the postural reactions of the skeletal muscles. The animal assumed a trance-like or cataleptic state in which muscular tone was retained, the general attitude being one of alertness. It stood with wide-open eyes but was quite unresponsive to all ordinary forms of stimulation. Pavlov looks upon this state as a transition stage between wakefulness and deep sleep, due to a less widespread diffusion of the inhibitory process, and similar in nature to hypnosis. It is suggested that during sleep, also, all cortical areas, analyzers as Pavlov terms them, are not necessarily under the inhibitory influence. The alertness of a mother, apparently in deep sleep, to the slightest noise made by her baby, is a case in point. In such an instance it is the auditory analyzer, or a part of it, which has remained apparently outside the inhibitory influence.

These observations have undoubtedly some bearing upon the production of *dreams*. Dreams are evidently due to cortical activity especially of the temporal lobes.

³ The electroencephalogram recorded from the cortex of the cat after transection at the lowermost limit of the medulla is essentially that of the normal waking state, after section at the upper border of the pons the electroencephalographic pattern is that characteristic of sleep. From this it could be inferred that a waking center exists somewhere in the brain stem between the two sections.

(ch. 68), since they involve vivid memory, and the ability to recall and bring together various sensory impressions. It would seem, however, that the activity of other regions possessing a more critical ability—those areas endowed with a higher analytic and synthesizing capacity—is in abeyance. The illogical, uncritical and often grotesquely absurd character of dreams is well known, yet the mental pictures are often drawn with great vividness, and a stimulus which in the waking state would leave little imprint upon consciousness is sometimes magnified enormously during sleep. As Descartes says, "A flea bit me and I dreamt of a sword cut!" Sensations arising in the viscera, such as those due to hunger, a distended stomach or bladder, thirst, etc., which during the day may cause no more than a passing thought may give rise to dreams filled with the most exciting events. It has been mentioned that dreams occur only during light sleep, that is, at a time when one would expect the internal inhibitory process to be restricted to the more highly specialized parts of the analyzers and before it has spread to involve the entire cortex. Dreaming may therefore be reasonably looked upon as being dependent upon a state of partial sleep—certain areas of the cortex being freed from the restraint which during the day is exercised by more critical regions.

(6) **KLEITMAN'S THEORY** According to Kleitman, sleep is due to the inactivity of the cerebral cortex resulting from a reduction in the number of afferent impulses, especially from the muscles, reaching the sensorium. Fatigue of the neuromuscular mechanism mediating muscle tone, with consequent suppression of impulses from muscle proprioceptors, is considered to be the most important factor in the onset of sleep. According to this author loss of muscle tone is an invariable prelude to sleep.

Kleitman claims to have demonstrated a diurnal variation in the speed, accuracy and steadiness with which certain muscular acts are performed. The efficiency of performance was maximal in the afternoon and minimal late at night and in the early morning. It is suggested that the variability is due to a corresponding rhythm in the tone of the skeletal musculature. But quite apart from these experiments, many observations give credence to the theory that cortical inactivity resulting from the blockage of afferent impulses is an important element in the onset of sleep. The exclusion of stimuli from visual, auditory and cutaneous receptors and the diminution of the flow of proprioceptor impulses as a result of muscular relaxation, are well-known means employed to induce sleep, whereas cortical activity, whether from psychic

causes—*anxiety, worry, excitement, etc.*—or as a result of impulses set up in exteroceptors or muscle receptors, prevents sleep. In *extreme fatigue* when muscle tone is presumably at a minimum and the threshold of other afferent paths also probably raised as well, sleep comes on irresistibly. Kleitman and his associates found that after a prolonged period of wakefulness the only way in which they could keep from falling asleep was by moving about, or at least remaining in a standing or sitting position. Upon lying down and permitting their muscles to relax they were immediately overpowered by sleep. Jacobson found that persons whom he had taught certain procedures for inducing progressive muscular relaxation fell asleep while they were relaxing.

An attempt has been made by Kleitman to reconcile the cortical theories of sleep with the undoubted fact that the hypothalamus is in some way concerned in the sleep process.

That sleep is not solely dependent upon the cortex is evident from the fact that decorticated dogs show periods of sleep alternating with intervals of wakefulness. The sleep rhythm in these animals is not, however, related in any way with night and day but consists of a number of shorter or longer periods throughout the twenty-four hours. The sleep periods occur most constantly after feeding. During the waking periods the animals walk around almost incessantly. Lower orders in the animal scale in which the cortex is rudimentary and which, in consequence, are unable to develop a wide variety of conditioned responses show a similar sleep rhythm. This more primitive sleep mechanism is dependent conceivably upon a center located at a subcortical level, and most probably in the hypothalamus. We may regard it as presiding over vegetative functions and acting continuously to keep the animal asleep unless inhibited by impulses arising out of the more primitive processes and reactions, e.g., hunger, thirst, cold or distension of the bladder or rectum.

The diurnal sleep rhythm, that is, the ability to keep awake throughout the day, is dependent, on the other hand, upon the development of conditioned reflexes. So long as the cortex can bring its analyzing ability to bear upon the stream of impulses received from the different distance receptors, muscles, skin, etc., the functions of the primitive sleep center are held in abeyance. With

the spread of internal inhibition over the cortex, or as a result of the elimination of stimuli from the periphery, the center asserts itself, and the subject is unable to remain awake. The diurnal sleep rhythm is therefore an acquired phenomenon—not inborn. Infants and very young animals do not show it but, like decorticated animals, have several sleeping periods throughout the day.

Those who believe that sleep results from the in-

hibition of a waking center in the hypothalamus take a somewhat different view. According to them sleep is a negative rather than a positive state, that is, it is due to the inactivity of the waking center. The activity of this center it is conceived is maintained by impulses received from the cortex. Cortical function is dependent in turn upon a flow of impulses along various afferent channels.

CHAPTER 70

THE CEREBELLUM

GENERAL STRUCTURE AND DIVISIONS

The cerebellum consists of a narrow central body, the *vermis* (or worm) and two lateral masses, the *right* and *left cerebellar hemispheres*. On its upper surface the demarcation between the vermis (*superior vermis*) and the hemispheres is slight. Upon the under surface the hemispheres are separated by a deep depression—the *vallecula*, the floor of the latter is formed by the inferior surface of the vermis. The inferior aspect of the vermis (*inferior vermis*) consists of four subdivisions, these are called in order from before backwards the *nodule*, *uvule*, *pyramid* and *tuber*. On either side and continuous with the nodule, is an elongated, somewhat lobulated structure called the *flocculus*.

The cerebellar surface is not convoluted like the cerebral cortex but is divided by parallel and curved furrows into numerous laminae or folia (leaves). The total cortical area of the human cerebellum is about 100,000 sq mm, or less than half that of the cerebral cortex.

Though the division of the cerebellum into the vermis and two hemispheres possesses considerable descriptive value, comparative neurologists (chiefly Bolk, Ingvar, Elliott Smith and Larsell), have suggested other divisions which possess greater significance from a phylogenetic and functional point of view. In Larsell's description the cerebellum is divided into two fundamental or primary parts, (a) the small *flocculonodular lobe* or *vestibular part*, and (b) the *corpus cerebelli* which is predominantly concerned with the integration of proprioceptive impulses (from the muscles). These two parts are separated by a deep fissure—the *posterolateral fissure*—which is present in all vertebrate brains.

The *flocculonodular lobe*, the most ancient part of the cerebellum, comprises the *flocculus* and the *nodule*. It is developed from the structures in the region of the vestibular nuclei. The *corpus cerebelli*, which includes the rest of the cerebellum, is separated from the *flocculonodular lobe* by the *fissura posterolateralis*, and is divided by a well marked fissure—the *fissura prima* of Elliott Smith—into a small anterior and a large posterior lobe. The *anterior lobe* consists of three subdivisions, the *lingula*, *lobulus centralis* and *calmen*. The *posterior lobe* includes the *lobulus simplex*, *declive*, *tuber*, *pyramid* and *uvule*, together with the associated parts of the hemispheres (*lobulus ansiformis* and *lobulus paramedianus*, see below) and the *paraflocculus* (fig 70.1). In higher forms a fissure appears in front of the pyramid, known as the *sulcus* or *fissura prepyramidalis*. The part of the posterior lobe between this sulcus behind and the *fissura prima* in front is sometimes referred to as the

middle lobe of Ingvar (fig 70.1), from the functional view this is a convenient subdivision. The *flocculonodular lobe*, the *anterior lobe*, and the *lobulus simplex*, *pyramid uvule* and *paraflocculus* of the posterior lobe are the phylogenetically old parts of the cerebellum and are referred to as the *paleocerebellum*. The remainder of the cerebellum, i.e., the lateral expansions or hemispheres (*ansiform lobules*) *declive* and *tuber* (*superior vermis*) are late acquisitions, and constitute the *neocerebellum*, they correspond to most of Ingvar's middle lobe, and appear in phylogenetic development at about the same time as the cerebral cortex and pyramidal tracts, and the pons. The *neocerebellum*, *cerebral cortex* and *pons* are absent or rudimentary in submammalian forms.¹

The *paramedian lobule* (or *tonsil*) is a small compact mass lying on either side of the inferior vermis. The *ansiform lobule*, which constitutes a large proportion of the posterior lobe and forms the expanded lateral mass of the hemisphere, reaches its greatest development in the human brain, its function is concerned with the tonus adjustments required in the performance of skilled muscular movements.

INTERNAL STRUCTURE

When sectioned in the sagittal plane each hemisphere of the cerebellum presents a branching core of white matter which from its foliage-like appearance, has been named the *arbor vitae*. The terminal branches of the white matter are covered with a coating of gray substance which constitutes the cerebellar cortex. The leaf-like structures so formed are spoken of as folia, and are responsible for the laminated appearance of the cerebellar surface (fig 70.2). Unlike the cortex of the cerebrum all areas of the cerebellar cortex show a uniform histological structure.

The gray matter

THE CORTEX Three cell layers are distinguished.

(1) The molecular (or plexiform) layer is outermost and consists largely of unmyelinated nerve

¹ The *flocculonodular lobe*, being the oldest part of the cerebellum, is sometimes called the *archicerebellum*. The first appearance of a cerebellum, phylogenetically, is in the primitive fish *Petromyzon* (lamprey), and consists merely of a bridge of nervous tissue formed by an outgrowth from either side of the medulla which fuse in the midline over the 4th ventricle. This is the forerunner of the *flocculonodular lobe* of higher forms.

fibers derived from (a) the white substance, (b) the cells of the two underlying layers, and (c) the cells within this layer itself. The cells of the molecular layer are arranged in a deep and a super-

run transversely in relation to the long axis of the folium and arborize by means of collaterals around the bodies of several Purkinje cells. They are referred to as "basket" cells (see fig. 70.3)

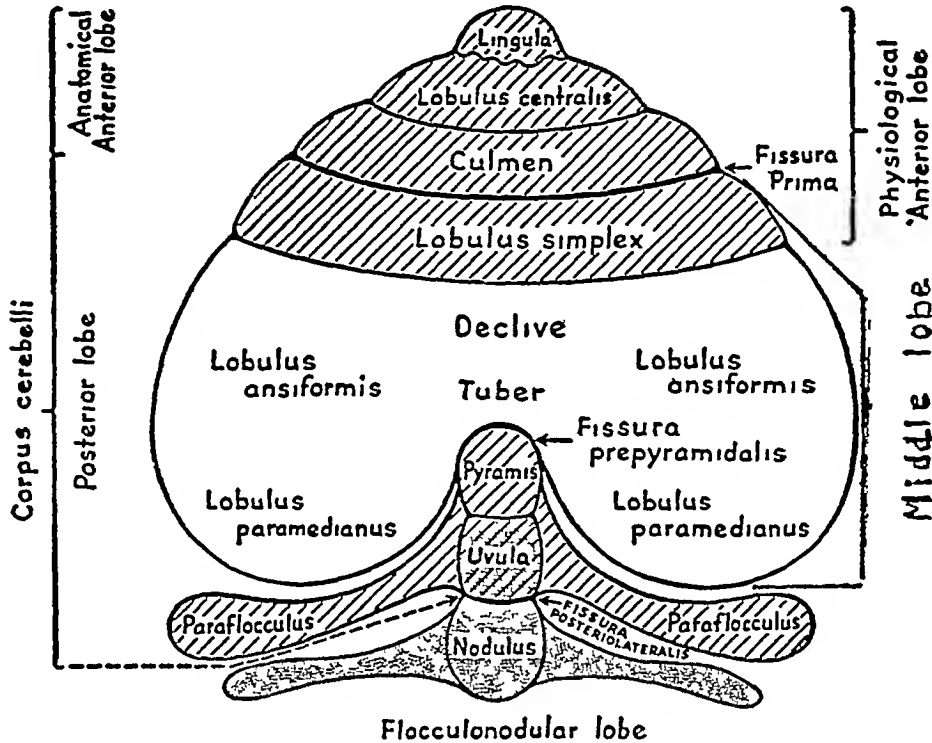


FIG 70.1 Divisions of the human cerebellum (After Larsell, modified). Afferent fiber connections: *stippled*, vestibular paleocerebellum; *diagonals*, spinal paleocerebellum; *clear*, corticopontocerebellar, neocerebellum.

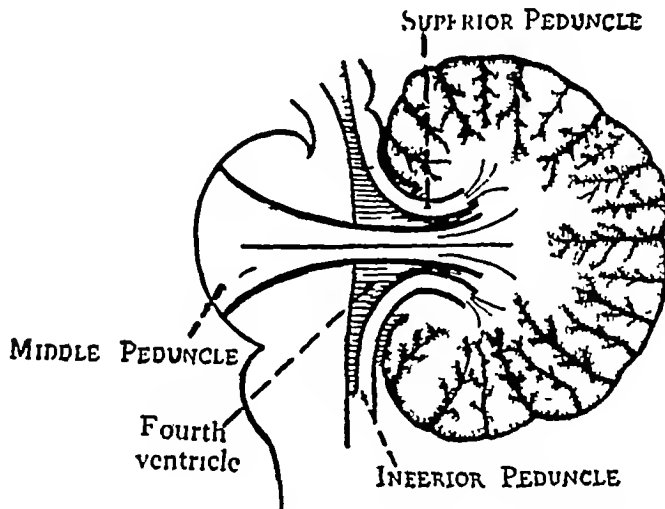


FIG 70.2 Diagram of the cerebellar peduncles (redrawn from Villiger)

ficial stratum, their axons synapse with the Purkinje cells, whose dendrites arborize throughout this layer. The cells of the superficial stratum are small, star-shaped and few in number. The deep stratum is composed of larger stellate cells whose axons

(2) The intermediate layer of Purkinje cells
The large flask-shaped bodies of these cells, which are peculiar to the cerebellum, form a layer between the molecular and the granular layers. Their dendrites pass outwards into the molecular layer

where they arborize luxuriantly, as just mentioned, collaterals of axons of the basket cells arborize around their bodies. The axons of the Purkinje cells enter the white substance, and end by synapsing with cells in the cerebellar nuclei. The Purkinje arborizations extend outwards through the entire thickness of the molecular layer. They are spread out or flattened in the transverse plane of the folium, thus resembling a vine trained against a wall rather than a bush. The body of the cell is flattened in a similar manner.

than the nucleus dentatus, which is found only in mammals. The fastigial nuclei, which are the most ancient of all, lie near the midline on either side in the roof of the 4th ventricle. They receive fibers from the paleocerebellum (anterior lobe, pyramid, uvula and flocculonodular lobe), and from the vestibular nuclei and 8th nerve through the inferior peduncle; they also project to the vestibular nuclei. The globose and emboliform nuclei, which are placed more laterally than the roof nuclei, receive fibers chiefly from the anterior lobe. The globose nucleus also receives fibers from

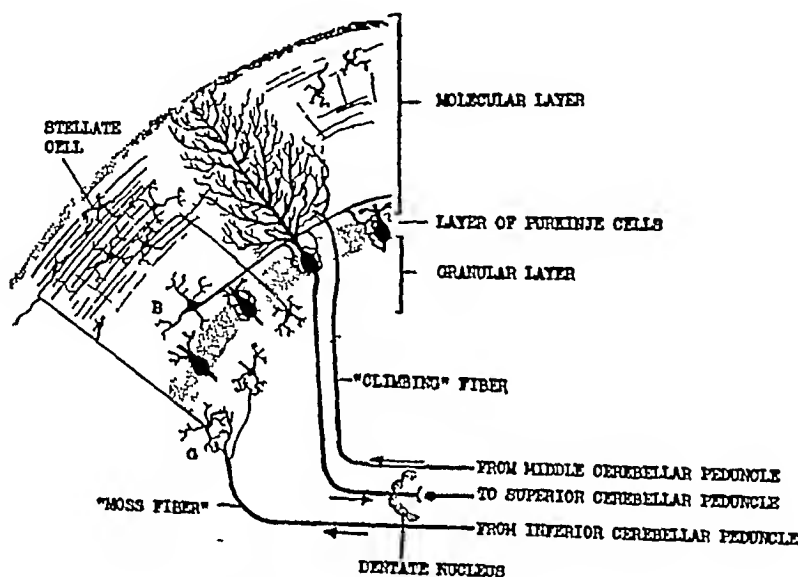


FIG 70.3 Diagram to show structure of cerebellar cortex G, granule cell, B, "basket" cell

(3) *The granular layer*, which rests upon the white matter, is composed of small, round, closely packed cells, and numerous nerve fibers. The cells possess four or five dendrites which end in a tuft of branches close to the cell body and connect with those of neighboring cells. One long process (axon) of each extends into the molecular layer where it connects with the dendrites of a large number of Purkinje cells. Afferent fibers ("moss fibers", p 1077) arriving via the inferior cerebellar peduncles make connections with the granule cells.

The cerebellar nuclei

The cerebellum contains on either side four separate gray masses. These are (a) *nucleus fastigi* (nucleus of the roof), (b) *nucleus globosus*, (c) *nucleus emboliformis*, and (d) *nucleus dentatus*. The first three of these are phylogenetically older

than the nucleus dentatus, which is found only in mammals. Both nuclei project through the superior peduncles to the large-celled nucleus of the red nucleus. The more lately acquired dentate nucleus, which is placed most laterally, is a large, crenated mass of gray matter, bent acutely upon itself. It received fibers from the neocerebellum, chiefly from the Purkinje cells of the ansiform lobule. It projects through the superior cerebellar peduncle to the small-celled nucleus of the red nucleus (and to some extent also to the large-celled nucleus), and to the ventrolateral group of the thalamic nuclei (fig 70.4).

The white matter, connections of the cerebellum with other parts of the central nervous system—cerebellar peduncles

The white matter of the hemispheres is composed of (a) *projection fibers*, i.e., fibers which

leave or enter the cerebellum via the peduncles, (b) *association fibers*, which connect different regions of the same hemisphere, and (c) *commissural fibers* connecting cortical areas of the two hemispheres

The flocculonodular lobe, mainly the nodule, is connected by both afferent and efferent fibers (relayed in the fastigial and globose nuclei) with the vestibular nuclei. The anterior lobes and the ansiform lobules receive spinocerebellar and pontocerebellar fibers, respectively, and project to brain

aroused by tactile stimuli to the cortex of the paramedian lobule of the posterior lobe and to the lobulus simplex of the anterior lobe

(3) *Vestibulocerebellar tract* from the vestibular nuclei of the same side, and also directly from the vestibular nerve. They pass to the three cerebellar nuclei (nucleus globosus, nucleus emboliformis and mainly to nucleus fastigi) and are relayed to the cortex of the flocculonodular lobe and of the uvule

(4) *Olivocerebellar tract* arising in the inferior olive of the opposite side and to some extent in the nucleus of the same side, the fibers of this tract end in the cor-

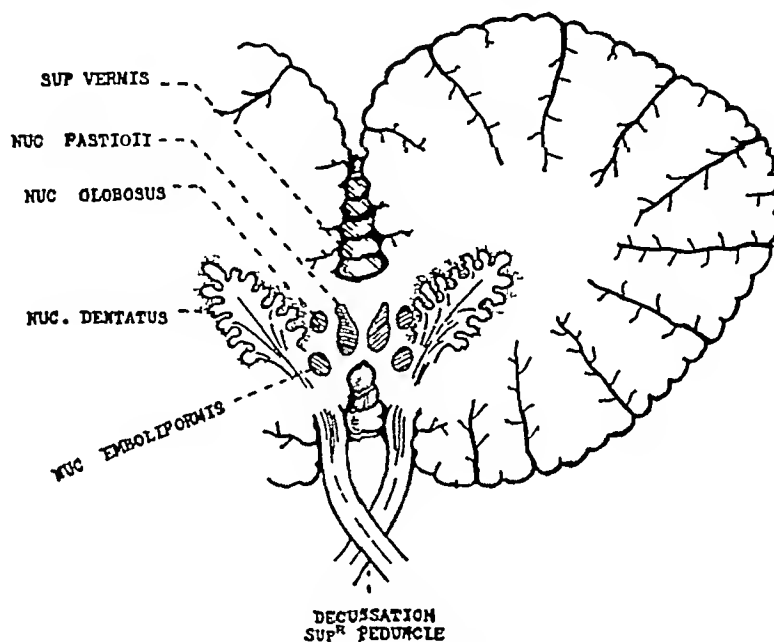


FIG 70 4 Diagram of a horizontal section through the cerebellum to show the cerebellar nuclei (viewed from in front)

stem and nuclei thalamus. The fibers, afferent and efferent, connecting the cerebellum with extracerebellar regions are all carried in three large bundles called the *inferior, middle and superior cerebellar peduncles*

THE INFERIOR PEDUNCLES (RESTIFORM BODIES) Its fibers are predominantly afferent and convey non-sensory impulses from the labyrinth, joints, voluntary muscles and skin (tactile). Its constituent fiber tracts are (figs 70 2 and 70 5)

A Afferent (entering) fibers

(1) *Dorsal (posterior) spinocerebellar (direct cerebellar) tract*. The fibers of this tract end in the cortex of the anterior and posterior lobes of both sides but mainly of the same side. Some fibers also end in the nodule

(2) *Dorsal external arcuate fibers* (p 990) from the nucleus gracilis and cuneatus of the same side, and the *ventral external arcuate fibers* from the corresponding nucleus of the opposite side. These fibers carry impulses

from those portions of the vermis and hemispheres constituting the posterior lobe

(5) Fibers of the 5th nerve and possibly of the 9th and 10th nerves which terminate in the pyramid, uvula and paraflocculus

(6) *Tectocerebellar*, from the colliculi to the cerebellum, the exact course and termination of this tract is unknown

B Efferent (leaving) fibers

(1) *Fastigiobulbar (or cerebellovestibular) tract*. This is a pathway from the flocculonodular lobe and the roof nucleus to the vestibular nucleus, and the reticular formation of the medulla. Impulses are relayed from the medulla via (i) the reticulospinal and the vestibulospinal tracts to the spinal centers, and (ii) the medial longitudinal fasciculus to the nuclei of the ocular nerves and into the anterior ground bundle of the cord

(2) *Cerebello-olivary tract* to the inferior olives of both sides. The latter are connected with the spinal centers through the olivospinal tracts (p 994)

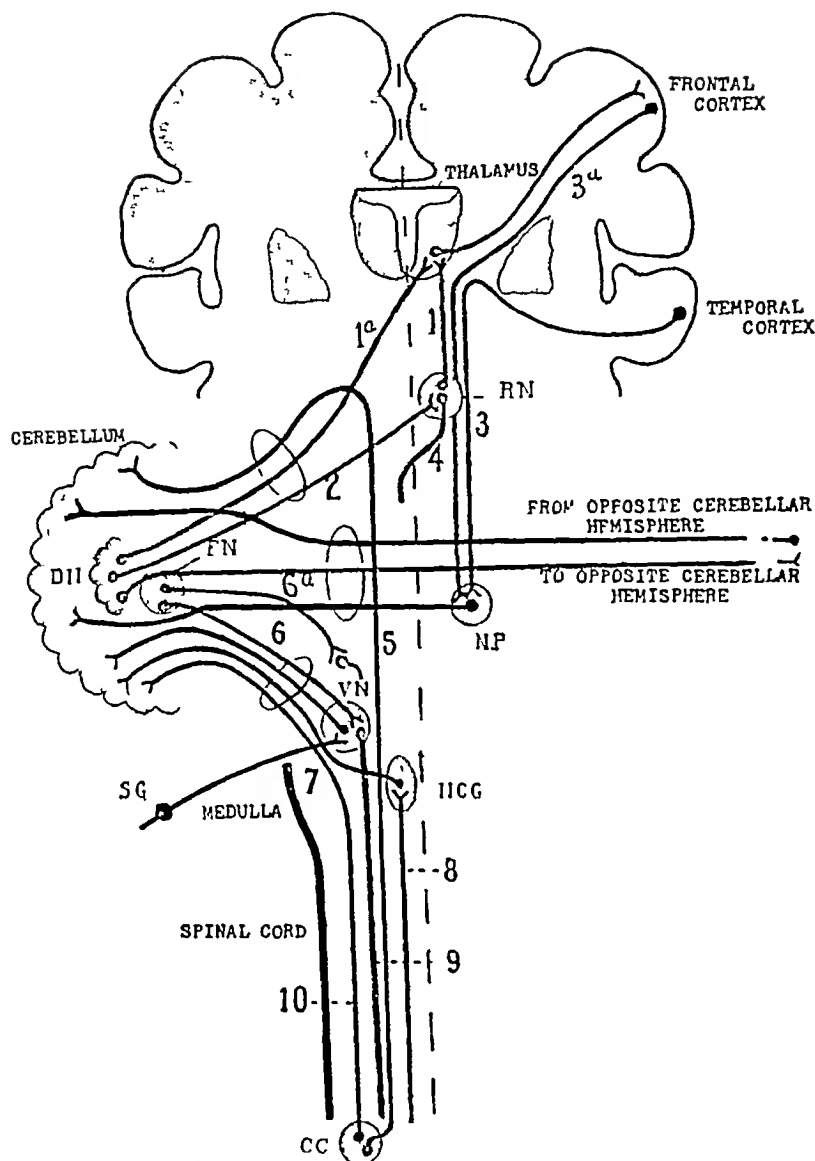


FIG 70.5 Diagram of chief cerebellar connections. Afferent fibers black, efferent fibers, red. RN, red nucleus, D.N., dentate nucleus, NF, fastigial nucleus, NP, pontine nuclei, VN, vestibular nucleus, N.G.C., nuclei gracilis and cuneatus, S.G., cell of Scarpa's ganglion, C.C., Clarke's column (dorsal nucleus). 1, dentatothalamic tract, 2, cerebellorubral tract, 3, temporo-pontine tract, 4, rubrospinal tract, 5, ventral (indirect) spinocerebellar tract, 6, fastigiovestibular tract, 6', fastigiobulbar tract, 7, external arcuate fibers passing from N.G.C. to cerebellum, 8, fasciculi gracilis and cuneatus, 9, vestibulospinal tract, 10, dorsal (direct) spinocerebellar tract.

THE MIDDLE PEDUNCLES (BRACHIA PONTIS) are also mainly afferent. Each contains

(a) fibers which arise from cells of the pontine nuclei and end in the cortex of the posterior cerebellar lobe (declive, tuber, and the ansiform and paramedian lobules) of the opposite side, but also in smaller numbers to the homolateral hemisphere and vermis. These fibers

constitute the secondary neurons of the *temporo-pontine-cerebellar* and the *fronto-pontine cerebellar* tracts (b) fibers which pass from the cerebellar nuclei to the hemisphere of the opposite side (see fig. 70.5).

THE SUPERIOR PEDUNCLES (BRACHIA CONJUNCTIVAE) enter into the formation of the upper part of the roof of the 4th ventricle and plunge into the

mid-brain just beneath the inferior colliculi. The superior cerebellar peduncle contains both efferent and afferent fibers but is composed predominantly of the former. It is through the superior peduncle that the cerebellum exerts its main influence upon the skeletal musculature.

A. The efferent fibers arise chiefly from the dentate nucleus, but a few are also derived from the nuclei globosus and emboliformis. The Purkinje cells of the cerebellar cortex constitute the primary neurons of these paths. The fibers decussate in the mid-brain with those of the opposite side and then divide into an ascending and a descending group.

(1) *The ascending fibers* pass (a) directly to the lateral ventral nucleus of the thalamus where the impulses are relayed to areas 4 and 6 of the cerebral cortex, (b) from the dentate nucleus to the small-celled nucleus (*n. parvocellularis*) of the red nucleus, and from the globose and emboliform nuclei to the large-celled nucleus (*n. magnocellularis*). The impulses reaching the small-celled nucleus (from dentate nucleus) are relayed via the rubro-thalamic tract to areas 4 and 6 of the cerebral cortex.

It will be noted that the ascending fibers connect one cerebellar hemisphere with the red nucleus, thalamus and cerebral cortex of the opposite side, but, as a result of the crossing of the rubrospinal and pyramidal (corticospinal and corticobulbar) tracts, each cerebellar hemisphere is ultimately connected with the same side of the brain stem and spinal cord.

(2) *The descending fibers* terminate around cells of the reticular formation of the pons, medulla and cervical cord.

B. The afferent fibers of the superior cerebral peduncle are

(1) *The ventral (indirect) spinocerebellar tract.* This ascends through the medulla and pons and reaching the upper level of the latter turns backwards, arches over the peduncle, enters the anterior medullary velum, and passes within this to the cerebellum. The fibers end in the cortex of the anterior lobe mainly of the same side.

(2) *The tectocerebellar tract*, composed of fibers which originate in the superior colliculus. It probably conveys retinal impulses and so constitutes a pathway for visuocerebellar reflexes.

Afferent-efferent (cerebro-cerebello-cerebral) circuit. This consists of fibers (fronto-pontine) from areas 4 and 6 of the cerebral cortex to the pontine nuclei, from the latter through the opposite middle cerebellar peduncle to the dentate nucleus, from this nucleus (described on p. 1074) via the superior peduncle to the brain stem where they cross to the red nucleus of the opposite side, thence to the posterior ventral nucleus of the thalamus, whence they return

to areas 4 and 6 of the cerebral cortex. Through these paths a steadying action is exerted upon voluntary movements initiated through the pyramidal tracts.

The fibers reaching the cerebellar cortex via the peduncles are of two main types: (a) "*moss fibers*," which end in moss-like appendages around the cells of the granular layer (fig. 70.3) and (b) *climbing fibers*, which pass outward to the molecular layer, here they give off collaterals resembling the tendrils of a vine which appear to cling to the arborizations of the Purkinje cells. According to Cajal the climbing fibers are derived from the vestibular and pontine nuclei, while the direct spinocerebellar and olivocerebellar tracts are constituted of moss fibers. Through their connections with the cells of the granular layer each moss fiber is connected indirectly with a large number of Purkinje cells, whereas a climbing fiber is in communication with only one or two Purkinje cells.

THE EXPERIMENTAL STUDY OF THE CEREBELLUM EFFECTS OF EXCISION OF THE CEREBELLUM AND OF SECTION OF THE PEDUNCLES

The earliest experiments upon the cerebellum were performed by Rolando (1809) and the French neurologist Flourens (1822). The latter removed the cerebellum of pigeons in which he observed, as a result of the ablation, grave disturbances of equilibrium and abnormal postures of the wings, neck and limbs. Luciani, in the later part of the 19th century, carried out cerebellar ablations upon dogs and observed three cardinal effects: (a) muscular weakness or *asthenia*, (b) a reduction in muscle tone, which he called *atonia* (really hypotonia), and (c) unsteadiness of voluntary movement, to which he gave the name *astasia*. These defects which appeared about a month after the operation gave rise to a coarse, jerky *tremor* upon attempting to perform any voluntary act.

Disequilibrium, due to the removal of the vestibular impulses, which is readily demonstrated in monkeys, causes a staggering, reeling, drunken-like gait—*cerebellar ataxia*. The animal—dog or monkey—stands with limbs spread in order to provide a broad base, sways from side to side with oscillations of the head.

Another characteristic symptom of cerebellar ablation, and which is also seen in lesions of the cerebellum in man, is *dysmetria*—the inability to gauge the strength or duration of the muscular contraction required to execute a certain voluntary act. For example, if the arm reaches for an object it may

overshoot the mark (*hypermetria*) or fall short of it (*hypometria*)

The deep reflexes are but little altered after decerebellation, though the knee jerk, rather than being a kick tends to be pendular in character. The postural reflexes (of antigravity muscles) on the whole are exaggerated as is also the positive supporting reaction. All the labyrinthine righting reflexes are retained, since their centers lie in the mid-brain. Decerebrate rigidity is enhanced, a fact first pointed out by Sherrington, and amply confirmed, destruction of the anterior lobe alone has this effect. Decerebellation leaves all types of sensation unaltered.

Unilateral ataxia. Removal of one half of the cerebellum causes a reduction in tone of the muscles of the same side and a tendency to fall to that side. The body may be curved toward the sound side.

Section of the peduncles. Section of all six peduncles causes effects identical with those of complete decerebellation, and division of the three of one side results in effects described for unilateral ablation. In the chimpanzee, *asyrerga*, *l. p'ora*, *easy fatigability*, and a *coarse tremor* appear on the same side as the section. Severance of one superior peduncle causes at first defects somewhat similar though less severe than those following unilateral section of all three peduncles, and is later compensated for through those which have been left intact. Little is known with respect to the disabilities following section of one or both middle peduncles. A lesion interrupting the fibers of the inferior peduncles causes, mainly, equilibrium disturbances (due to division of vestibular connections), and ataxia as a result of the interruption of spinocerebellar fibers. Compensation through the intact cerebellar half tends to occur later, so that the effects become progressively less severe.

Even after unilateral section of the three peduncles, compensation is brought about after a time through the remaining half of the cerebellum and through the frontal lobes (Botterell and Fulton).

The experiments of Aring and Fulton indicate that the nervous mechanism involved in the production of cerebellar tremor lies in the excitable part of the cerebral cortex (area 4 and the upper part of area 6), removal of these areas after contralateral section of the cerebellar peduncles abolished the tremor. The ability to compensate for cerebellar defects depends, apparently, upon the premotor area (area 6, upper part), removal of which is followed by marked accentuation of the cerebellar signs and permanent impairment of the animal's

ability to compensate for the cerebellar defect. These findings suggest that the "cerebellar" signs which are sometimes seen clinically in lesions of the frontal lobe are probably due to involvement of the premotor cerebellar connections (frontopontine-cerebellar tract). On the other hand, removal of the motor area (area 4) alone, temporarily abolishes and permanently depresses the manifestations of cerebellar deficiency.

STIMULATION OF THE CEREBELLAR NUCLEI

The following is a summary of the results obtained by Miller and Laughton

(a) Faradic stimulation of the *nucleus dentatus* causes increased tone of the flexor muscles of the ipsilateral limbs (biceps brachii and tibialis anticus) together with inhibition of the tone of the extensor muscles (lateral head of the triceps, gastrocnemius and soleus). There is increased tone of the trunk muscles of the stimulated side with consequent curvature of the body. There is no after discharge in the case of the limb flexors, "rebound" occurs in the extensors after cessation of the stimulation.

(b) Stimulation of the *nucleus emboliformis*, *globosus* or *fastigii* causes similar but more intense responses. Ocular movements may occur. In the case of the *nucleus fastigii* strong flexion of both forelimbs may result.

The efferent pathways for these reactions are via the superior peduncle, red nucleus and rubrospinal tracts in the case of the *nucleus dentatus*, *emboliformis* and *globosus*, and through the inferior peduncle, vestibular (Deiter's) nucleus and vestibulospinal tract in the case of the *nucleus fastigii*.

Localization in cerebellar cortex

In recent years representation in the cerebellum, afferent and efferent, have been shown to be much more circumscribed than had been supposed. Yet, even so, localization has not been found to be as precisely discrete as in certain regions of the cerebral cortex, e.g., in the motor area. Owing to the double crossing of the pathway for impulses from the cerebellum, as they ascend in the mid-brain (dentatorubral tract) and as they descend to the spinal cord (rubrospinal tract), each half of the cerebellum exerts its influence mainly on the musculature of the same side of the body.

In the decerebrate animal stimulation of the anterior lobe inhibits the extensor rigidity, but, as already mentioned, the latter is accentuated by excision of the anterior lobe. In normal animals hypotonia is caused by stimulation of the cortex of this part of the cerebellum.

Representation in the cerebellar cortex of somatic receptive areas has been studied by Dow, Snyder and his associates, and by Adrian, by means of the method of evoked potentials (ch 68) Adrian recorded the potentials evoked in the cortex of the *anterior lobe* of cats and monkeys by means of a fine wire electrode inserted beneath the surface. Various receptive areas over the limbs and face were stimulated by touch, pressure, joint movement, or stretching of muscles (fig 70 6)

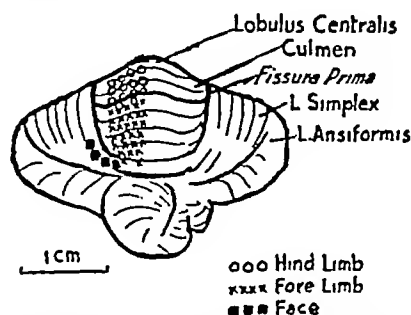


FIG 70 6 Spinocerebellar receiving areas for hind-limb, fore-limb and face (vibrissae) in the cat, as determined by the method of evoked potentials (From Adrian)

The representation in the anterior lobe of the different somatic areas is in reverse order. That is to say, the hindlimb area is placed anteriorly to the areas for the forelimbs and face. Thus, stimulation of the hindlimb evoked potentials from the lobulus centralis and an anterior strip of the culmen, stimulation of the forelimb induced responses from the remainder of the culmen, and stimulation of the face (vibrissae in cat) from the lobulus simplex. The spinocerebellar impulses from the tail are probably received by the most anteriorly situated part of the anterior lobe, namely, in the *lingula*.² An even, more precise localization was mapped out within these areas, in the area for the hindlimb, the foot is in front of the knee and the knee in front of the thigh, in the forelimb representation the shoulder in front of the wrist, with the wrist and hand following in corresponding order.

Potentials were also evoked from the anterior lobe by stimulation (electrical or local application of strychnine) of the motor area of the cerebral cortex, i.e., by causing a discharge over corticopontine and pontine-cerebellar pathways. The receiving areas of these impulses lie more laterally and extend into the ansiform lobules. The upper part of the motor area of the cerebral cortex, (i.e.,

² In the gaffe the lobulus simplex is greatly enlarged, and in monkeys with a prehensile tail the *lingula* is elongated.

area for hindlimb) was found to discharge to the contralateral hindlimb area of the anterior lobe (lobulus centralis), the lower part (face area) to the lobulus simplex, and the forelimb area of the precentral gyrus to the cerebellar area in between, namely the culmen (fig 70 7)

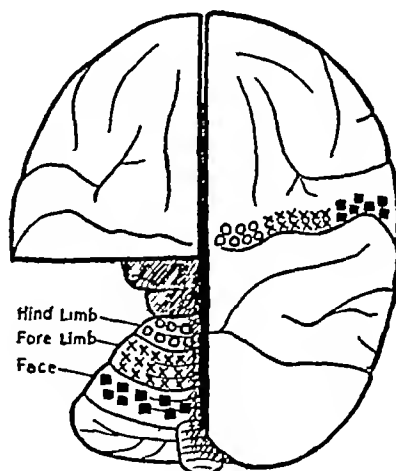


FIG 70 7 Subdivisions of pontocerebellar receiving area in the monkey, showing their connection with different parts of the motor cortex (From Adrian)

Impulses from the auditory and visual systems are also conveyed to the cerebellum. Sound, or flashes of light, evoke potentials from an area covering the lobulus simplex, declive and tuber, and extending laterally into the ansiform lobule. The visual impulses are relayed to the cerebellum from the superior colliculus, the auditory from the inferior colliculus. The tectocerebellar tracts probably provide the pathways for these impulses.

From ablation experiments topographical localization similar in character to that described above for the spinocerebellar and corticopontine-cerebellar systems has been demonstrated for the efferent side. The tail muscles are represented in the *lingula*, the muscles for the hindlimbs in the lobulus centralis, the forelimb muscles in the culmen, and the cervical and facial muscles in the lobulus simplex. Representation of the limbs in the middle lobe of Ingvar (ansiform lobule, declive and tuber) has not been demonstrated, but from the relatively enormous size of this part of the cerebellum in higher mammals and from the fact that it (ansiform lobule) is the chief receiving station for corticopontocerebellar fibers, it cannot be doubted that it is of paramount importance in volitional movement.

The *flocculonodular lobe*, through its vestibular connections, is concerned with equilibratory function, severe disturbances of balance resulting from

overshoot the mark (*hypermetria*) or fall short of it (*hypometria*)

The deep reflexes are but little altered after decerebellation, though the knee jerk, rather than being a kick tends to be pendular in character. The postural reflexes (of antigravity muscles) on the whole are exaggerated as is also the positive supporting reaction. All the labyrinthine righting reflexes are retained, since their centers lie in the mid brain. Decerebrate rigidity is enhanced, a fact first pointed out by Sherrington, and amply confirmed, destruction of the anterior lobe alone has this effect. Decerebellation leaves all types of sensation unaltered.

Unilateral ablation. Removal of one half of the cerebellum causes a reduction in tone of the muscles of the same side and a tendency to fall to that side. The body may be curved toward the sound side.

Section of the peduncles. Section of all six peduncles causes effects identical with those of complete decerebellation, and division of the three of one side results in effects described for unilateral ablation. In the chimpanzee, *asynergia*, *hypotonia*, *easy fatigability*, and a *coarse tremor* appear on the same side as the section. Severance of one superior peduncle causes at first defects somewhat similar though less severe than those following unilateral section of all three peduncles, and is later compensated for through those which have been left intact. Little is known with respect to the disabilities following section of one or both middle peduncles. A lesion interrupting the fibers of the inferior peduncles causes, mainly, equilibratory disturbances (due to division of vestibular connections), and ataxia as a result of the interruption of spinocerebellar fibers. Compensation through the intact cerebellar half tends to occur later, so that the effects become progressively less severe.

Even after unilateral section of the three peduncles, compensation is brought about after a time through the remaining half of the cerebellum and through the frontal lobes (Botterell and Fulton).

The experiments of Aring and Fulton indicate that the nervous mechanism involved in the production of cerebellar tremor lies in the excitable part of the cerebral cortex (area 4 and the upper part of area 6), removal of these areas after contralateral section of the cerebellar peduncles abolished the tremor. The ability to compensate for cerebellar defects depends, apparently, upon the premotor area (area 6, upper part), removal of which is followed by marked accentuation of the cerebellar signs and permanent impairment of the animal's

ability to compensate for the cerebellar defect. These findings suggest that the "cerebellar" signs which are sometimes seen clinically in lesions of the frontal lobe are probably due to involvement of the premotor cerebellar connections (frontopontine-cerebellar tract). On the other hand, removal of the motor area (area 4) alone, temporarily abolishes and permanently depresses the manifestations of cerebellar deficiency.

STIMULATION OF THE CEREBELLAR NUCLEI

The following is a summary of the results obtained by Miller and Laughlin.

(a) Faradic stimulation of the *nucleus dentatus* causes increased tone of the flexor muscles of the ipsilateral limbs (biceps brachii and tibialis anticus) together with inhibition of the tone of the extensor muscles (lateral head of the triceps, gastrocnemius and soleus). There is increased tone of the trunk muscles of the stimulated side with consequent curvature of the body. There is no after discharge in the case of the limb flexors, "rebound" occurs in the extensors after cessation of the stimulation.

(b) Stimulation of the *nucleus emboliformis*, *globosus* or *fastigii* causes similar but more intense responses. Ocular movements may occur. In the case of the *nucleus fastigii* strong flexion of both forelimbs may result.

The efferent pathways for these reactions are via the superior peduncle, red nucleus and rubrospinal tracts in the case of the *nuclei dentatus*, *emboliformis* and *globosus*, and through the inferior peduncle, vestibular (Deiter's) nucleus and vestibulospinal tract in the case of the *nucleus fastigii*.

Localization in cerebellar cortex

In recent years representation in the cerebellum, afferent and efferent, have been shown to be much more circumscribed than had been supposed. Yet, even so, localization has not been found to be as precisely discrete as in certain regions of the cerebral cortex, e.g., in the motor area. Owing to the double crossing of the pathway for impulses from the cerebellum, as they ascend in the mid brain (dentatorubral tract) and as they descend to the spinal cord (rubrospinal tract), each half of the cerebellum exerts its influence mainly on the musculature of the same side of the body.

In the decerebrate animal stimulation of the anterior lobe inhibits the extensor rigidity, but, as already mentioned, the latter is accentuated by excision of the anterior lobe. In normal animals hypotonia is caused by stimulation of the cortex of this part of the cerebellum.

Representation in the cerebellar cortex of somatic receptive areas has been studied by Dow, Snyder and his associates, and by Adrian, by means of the method of evoked potentials (ch 68) Adrian recorded the potentials evoked in the cortex of the *anterior lobe* of cats and monkeys by means of a fine wire electrode inserted beneath the surface. Various receptive areas over the limbs and face were stimulated by touch, pressure, joint movement, or stretching of muscles (fig 70 6)

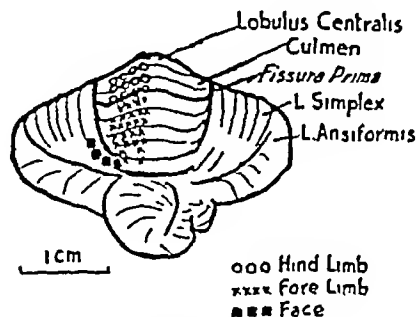


FIG 70 6 Spinocerebellar receiving areas for hindlimb, fore-limb and face (vibrissae) in the cat, as determined by the method of evoked potentials (From Adrian)

The representation in the anterior lobe of the different somatic areas is in reverse order. That is to say, the hindlimb area is placed anteriorly to the areas for the forelimbs and face. Thus, stimulation of the hindlimb evoked potentials from the lobulus centralis and an anterior strip of the culmen, stimulation of the forelimb induced responses from the remainder of the culmen, and stimulation of the face (vibrissae in cat) from the lobulus simplex. The spinocerebellar impulses from the tail are probably received by the most anteriorly situated part of the anterior lobe, namely, in the lingula.² An even, more precise localization was mapped out within these areas, in the area for the hindlimb, the foot is in front of the knee and the knee in front of the thigh, in the forelimb representation the shoulder in front of the wrist, with the wrist and hand following in corresponding order.

Potentials were also evoked from the anterior lobe by stimulation (electrical or local application of strychnine) of the motor area of the cerebral cortex, i.e., by causing a discharge over corticopontine and pontine-cerebellar pathways. The receiving areas of these impulses lie more laterally and extend into the ansiform lobules. The upper part of the motor area of the cerebral cortex, (i.e.,

² In the giraffe the lobulus simplex is greatly enlarged, and in monkeys with a prehensile tail the lingula is elongated.

area for hindlimb) was found to discharge to the contralateral hindlimb area of the anterior lobe (lobulus centralis), the lower part (face area) to the lobulus simplex, and the forelimb area of the precentral gyrus to the cerebellar area in between, namely the culmen (fig 70 7)

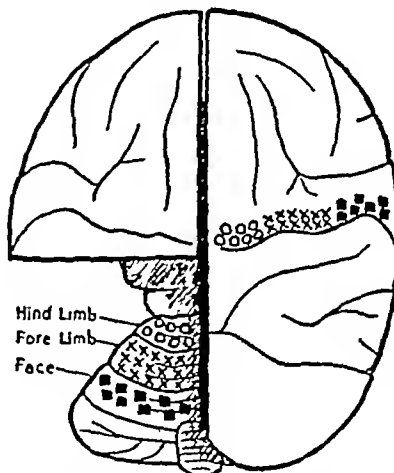


FIG 70 7 Subdivisions of pontocerebellar receiving area in the monkey, showing their connection with different parts of the motor cortex (From Adrian)

Impulses from the auditory and visual systems are also conveyed to the cerebellum. Sound, or flashes of light, evoke potentials from an area covering the lobulus simplex, declive and tuber, and extending laterally into the ansiform lobule. The visual impulses are relayed to the cerebellum from the superior colliculus, the auditory from the inferior colliculus. The tectocerebellar tracts probably provide the pathways for these impulses.

From ablation experiments topographical localization similar in character to that described above for the spinocerebellar and corticopontine-cerebellar systems has been demonstrated for the efferent side. The tail muscles are represented in the lingula, the muscles for the hindlimbs in the lobulus centralis, the forelimb muscles in the culmen, and the cervical and facial muscles in the lobulus simplex. Representation of the limbs in the middle lobe of Ingvar (ansiform lobule, declive and tuber) has not been demonstrated, but from the relatively enormous size of this part of the cerebellum in higher mammals and from the fact that it (ansiform lobule) is the chief receiving station for corticopontocerebellar fibers, it cannot be doubted that it is of paramount importance in volitional movement.

The *flocculonodular lobe*, through its vestibular connections, is concerned with equilibratory function, severe disturbances of balance resulting from

its ablation. Ablation of the nodule in dogs confers complete immunity to motion sickness (Bard and associates, see page 574). Removal of no other part of the cerebellum or of the cerebral cortex has this protective action (ch 65). Ablation of the *nodule* and *flocculus* (flocculonodular lobe, p 1072) in monkeys or of the nodule and part of the uvule (Dow) which causes injury to vestibular paths, is followed by pronounced disturbances of balance. In man a lesion of this part of the vermis is also associated with disturbances of equilibrium, e.g., "forced" falling. Removal of the *nodule* alone in monkeys is followed by oscillation of the head and neck (and, in some instances of the whole trunk) forced falling backwards and an abducted gait.

The functions of the *uvule* are largely obscure, its destruction alone is followed by some transient disturbance of equilibrium. Very little is known of the function of the *pyramid*, though it appears to be concerned in some way with vision or eye movement. Its excision causes no defect of balance. Ablation of the pyramid in the monkey results in an inability to halt its forward progression in time to prevent it from crashing into some obstacle. Faradic stimulation of the pyramid elicits an upward movement of the eyes, and Fulton suggests from this fact and the preceding observation the possibility of the pyramid being concerned in some way with the integration of proprioceptive impulses essential in judging distance.

Summary of functional localization

(1) All cerebellar activities, afferent and efferent, are carried out beneath consciousness.

(2) In the anterior lobe are represented in regular order the various receptive and muscular regions of the body. It is concerned mainly with the adjustments of muscle tone and synergic action necessary for posture. It receives spinocerebellar fibers (ventral and dorsal spinocerebellar tracts, and dorsal external arcuate fibers carrying impulses from the nuclei gracilis and cuneatus of the medulla).

(3) The middle lobe of Ingvar (neocerebellum) is the chief receptive region for corticopontineocerebellar tracts. It is most highly developed in higher mammals.

(4) The flocculonodular lobe, the most ancient part of the cerebellum, is intimately connected with the vestibular apparatus both through the vestibular nuclei and directly by fibers of the vestibular nerve, and therefore constitutes an important part of the neural mechanisms underlying equilibratory function.

(5) The functions of the uvule and pyramid are little known, though the latter probably has to do with eye movements and the former, to a minor degree, with equilibrium.

(6) Viewed in a broad and more general way we see the cerebellum as an organ to which impulses from proprioceptors in muscles, joints and vestibules, exteroceptors of the body surface as well as from auditory and visual receptors.

Electrical stimulation of the cortex of the hemispheres causes no obvious excitatory or inhibitory effect, though a rise in excitability of the motor area of the cerebral cortex can be induced by this means, and Miller found that the application of a 1 per cent solution of strychnine to the surface of the cerebellar hemisphere caused increased extensor tone and clonic movements of the limbs, chiefly of the same side. The effects following unilateral excision of one hemisphere (monkey) or its decortication are surprisingly slight, consisting of *hypotonia*, some difficulty in the execution of skilled muscular acts and disturbances of gait (Botterell and Fulton). These defects involve both limbs on the same side as the lesion and are transient, persisting for only a few days (1 to 2 weeks).

GENERAL SUMMARY OF CEREBELLAR FUNCTION

Thus, the cerebellum is, as Sherrington has called it, "the head ganglion of the proprioceptive system". Through its afferent connections the cerebellum is kept in constant touch with all parts of the nervous system having to do with voluntary movements, and those postural adjustments of a reflex or semi-automatic nature which are required for the maintenance of equilibrium. Through its efferent connections especially with areas 4 and 6 of the cerebral cortex it brings an influence to bear upon the various muscle groups engaged in a voluntary movement which enables them to act smoothly and steadily as a cooperative whole, similarly, it exerts a steadying effect upon the continuous muscular action required for the maintenance of posture, e.g., holding the head erect, and the trunk poised upon the limbs.

Following Beever, the muscles concerned in a voluntary muscular act may be classed in four groups.

(a) **PRIME MOVERS OR AGONISTS**, those whose contraction is essentially responsible for the movement of the part.

(b) **ANTAGONISTS**, those which oppose the prime movers.

(c) **SYNERGISTS**, those which assist the prime movers and reduce unnecessary movements to a minimum

(d) **FIXATION MUSCLES**, those whose contraction causes the fixation of the neighboring joints and maintains the limb or body in a position appropriate for carrying out the particular movement

Normally these muscle groups act as a unit. The contraction of the prime mover, e.g., the biceps, is accompanied by inhibition of the contraction of its antagonist, i.e., the triceps. The reciprocal inhibition, however, is not absolute, the antagonistic muscle as shown by Tilney and Pike, gives a coincident but weaker contraction, and thus serves the function of a "brake" or "snubber" to check the action of the agonist. At the same time the fixation muscles of the neighboring joint contract, thus enabling the agonist to exert its force with a minimum of waste effort. Any increase in the force of contraction of the agonist is accompanied by a corresponding rise in tone of its opponent (antagonist). Also, the muscles of fixation and the synergists contract more powerfully as the contraction of the prime movers increases, while groups of fixation muscles and synergists, which in a weaker movement remain quiescent, may be called into play. To give an example, when the fingers are closed with moderate force the only fixation muscles which show activity are the extensors of the wrist. When, however, the fist is clenched forcibly, muscles of the elbow and even those of the shoulder contract to fix these joints. Through such action between the various muscles taking part in a given act, smoothness and steadiness of movement are ensured and muscular force thereby economized. After destruction of the cerebellum these nice coöperative actions are lost. The asynergia results in jerkiness, overaction and imperfect muscular control (ataxia).³ The force of a given muscular movement is not graded with precision for the work to be done, but "over-shoots the mark." This led Babinski to suggest originally, that cerebellar activity (through the cerebral cortex) served as a "brake" to check such a tendency, that is, to overcome the inertia of the moving part.

³ Holmes denies that the cerebellum exerts any specific function in respect to the distribution of tone among various muscles acting synergically. The steady-ing action which tonic muscles normally exert over their antagonists is absent in cerebellar disease but Holmes maintains that this is due merely to the general hypotonicity.

THE MANIFESTATIONS OF CEREBELLAR DISEASE

The signs of cerebellar disease are much more pronounced in acute lesions, e.g., abscess, hemorrhage or trauma, than in those, e.g., tumor, which develop more gradually. In slowly developing lesions or after the subsidence of acute cerebellar disease a certain degree of compensation for the cerebellar defect always occurs. The majority of the signs of cerebellar disease are the result fundamentally of asynergia of the voluntary muscles.

Following Holmes, the chief features of cerebellar lesions will be considered under the following headings: (1) *hypotonia*, (2) *asthenia*, *fatigability* and *slowness of movement*, (3) *tremor*, (4) *asynergia*, (5) *decomposition of movement*, (6) *dysmetria*, (7) *rebound phenomenon*, (8) *adiadochokinesis*, (9) *vertigo and past-pointing*, (10) *deviation of the eyes and nystagmus*, (11) *attitude and gait*, (12) *disturbances of speech*, (13) *reflexes*.

HYPOTONIA This is usually a prominent feature in acute cerebellar lesions. The muscles are flabby. The limbs for this reason assume unnatural attitudes and can be moved passively into positions of extreme flexion or extension. The extremities flop about when shaken vigorously and, if the patient is suddenly rotated when standing, the arms swing loosely from the shoulders.

ASTHENIA, FATIGABILITY AND SLOWNESS OF MOVEMENT The muscles on the affected side are definitely weakened and tire easily. The commencement of any voluntary movement is delayed, and both contraction and relaxation phases are abnormally slow. Neither hypotonia nor asthenia are notable features in chronic lesions and it is probable that in acute lesions hypotonia is not a result of the cerebellar defect itself but of the involvement of neighboring structures. It will be recalled that hypotonia was not observed by Magnus nor by Aring and Fulton after decerebellation. Some also question whether asthenia is a manifestation of cerebellar deficit, but Holmes considers it to be due to involvement of the cerebellar nuclei.

TREMOR This is a coarse involuntary jerking which supervenes upon any attempt at voluntary movement (*intention tremor*), or when the muscles are engaged in maintaining a steady posture (*static tremor*), as in holding the head erect. The pyramidal tracts are primarily responsible for the tremor, but normally inhibitory impulses arising in areas 4 and 6, and travelling in the cerebro cerebello-cerebral circuit mentioned on page 1077 exert a steadying

movement upon voluntary movement, which is lost when this circuit is interrupted in any part of its course. The tremor is abolished in man by sectioning the lateral pyramidal tract (Putnam), as already mentioned. Removal of areas 6 and 4 abolishes cerebellar tremor in the monkey.

ASYNERGIA It should be pointed out that this term as used by Holmes has a much more restricted meaning than that which is usually given to it. This observer applied the term especially to the asynchronism observed between the actions of the fixation muscles and the prime movers, as well as to the lack of coöperation which normally exists between the muscles of the limbs, trunk and neck during walking and standing, and other acts necessitating postural adjustments of the various members of the body against gravity. For example, when the patient is asked to clench his fist, the extensors which normally fix the wrist as the fingers are flexed, either contract too early, and so over-extend the joint, or too late. In the latter event the wrist is flexed as the fist closes. Also, the patient may fall when he throws his head back owing to his failure to flex his knees, and so maintain his center of gravity over his base. Similarly when he attempts to sit upon a low stool he may fall backwards because he cannot at the same time flex his trunk at the hips.

DECOMPOSITION OF MOVEMENT The patient performs acts "by numbers", to use a military expression. When, for example, he is asked, as he lies in bed, to extend his arm vertically, and then to touch his nose with his forefinger, the latter movement is not performed naturally by simultaneously lowering the arm and bending the elbow. The extended arm is first lowered to the side, not until this position is reached is the elbow flexed and the finger brought to the nose.

DYSMETRIA is the lack of ability to adjust the force of the contraction necessary for the accomplishment of a given act. For example, when asked to touch a point with his finger the patient overshoots the mark or, less commonly, fails to reach it. Dysmetria in a lower limb may be demonstrated by the heel to knee test, when the patient, while recumbent, is directed to rest the heel of the affected side upon the opposite knee, he brings it in contact with the thigh or with the leg below the knee. Dysmetria is another manifestation of the general lack of synergic muscular control.

REBOUND PHENOMENON If the patient is asked to attempt a movement against a resistance which

is then suddenly removed, the limb moves forcibly in the direction towards which the effort was made. For example, if the observer holds the wrist of the patient, then asks him to bend the elbow, and, while he is making the effort, the forearm is suddenly released, flexion of the arm occurs with an unusual degree of force. This rebound phenomenon can be attributed to the absence of the "braking" action (p. 1081) of the antagonistic muscles.

ADIADOCHOKINESIS (a = primitive, diadocho = succession, kinesis = movement) is the name given by Babinski to the inability to execute alternating movements rapidly, e.g., pronation and supination of the forearm, or flexion and extension of the fingers.

VERTIGO AND PAST-POINTING (p. 982) When the patient with his eyes closed is asked to raise the arm of the affected side and to touch a prescribed mark, the arm deviates outwards, rarely inwards. When directed to move the finger in the horizontal plane and touch the mark the arm in some cases, deviates upwards, in others downwards.

DEVIATION OF THE EYES AND NYSTAGMUS In an acute unilateral lesion there may be conjugate deviation of the eyes to the uninjured side. When the patient is asked to look at an object in front of him the eyes move into position but then slowly deviate again to the uninjured side, after full deviation the eyes may return again to the central position with a sharp jerk. These alternate horizontal movements occur repeatedly and are known as nystagmus (p. 979). Though no deviation of the eyes is evident, nystagmus can sometimes be demonstrated in the subject of a cerebellar lesion when he is asked to turn his eyes towards either the injured or the uninjured side, the eyes are moved in the direction indicated and then deviate slowly towards the mid line, only to be returned in a series of jerks toward the point in which he is attempting to look. The nystagmus is due to injury of vestibulo-cerebellar paths (see also p. 984).

Skew deviation of the eyes is sometimes observed in acute cerebellar disease.

ATTITUDE AND GAIT Cerebellar disease sometimes results in abnormal attitudes. The trunk may be concave towards the affected side and sometimes rotated with the affected shoulder advanced. Rotation of the head with the chin towards the sound side, and flexed towards the shoulder of the affected side is not uncommonly seen in

unilateral cerebellar disease (This attitude is probably due to an associated lesion of the vestibular paths in the brain stem) In standing the body is inclined towards the side of the lesion and tends to fall to that side Falling backwards or forwards may occur (especially in lesions of the vermis) The abnormal attitudes of the limbs resulting from the muscular hypotonicity when this is present, have been mentioned

The gait is often staggering, reeling or lurching in character The patient's line of travel is not straight but tends to be curved, the deviation is towards the affected side When he is able to see his way he attempts to correct his tendency to deviate to one side by bringing himself back from time to time to his intended line of travel Thus he follows a zig-zag course to his objective. The movements of the lower limbs are ataxic, they are often thrown about in an awkward, uncontrolled manner, the feet are raised unnecessarily high and brought down again clumsily, often in a clapping fashion

DISTURBANCES OF SPEECH are due to asynergia of the muscles of phonation and articulation It is seen most frequently in lesions of the vermis The speech abnormality may be drawing, scanning, sing-song or explosive in type.

THE REFLEXES The cutaneous reflexes are normal and the deep reflexes, with the exception of the knee jerk, show as a rule no marked departure from the normal The knee jerk is not uncommonly pendular in character That is to say,

when the leg is hanging free a tap on the patellar tendon causes a slower, less brisk response than normal, but one of greater amplitude Also, unlike the response of the sound side, the return excursion of the leg is not arrested when, as a result of the influence of gravity, it reaches the resting position, but swings beyond Several to and fro movements follow before the limb finally comes to rest This behavior is evidently due to the hypotonicity of the flexor and extensor muscles and thus to a lack of the restraining effect which they normally exert upon one another to prevent exaggerated excursions of the limb in either direction

Muscle sense and other forms of sensation are unaffected in cerebellar disease The following are the principal types of cerebellar lesion (1) *tumor*, (2) *abscess*, (3) *injury*, e.g., gunshot wounds, (4) *degenerations* of (a) the cerebellar cortex, (b) the middle and inferior peduncles, (c) more rarely, of the cerebellar nuclei and superior peduncles, (d) of the spinocerebellar tracts, e.g., Friedreich's ataxia, a hereditary condition

A lesion involving the neocerebellum (superior vermis and hemispheres) is associated with hypotonia, dysmetria, weakness, slowness and irregularity of voluntary movement, tremor and nystagmus In disease of the flocculonodular lobe, there are marked disturbances of balance with a swaying, staggering gait and a tendency to fall backwards

CHAPTER 71

THE CEREBROSPINAL FLUID (C S P)

The discovery of the cerebrospinal fluid is generally ascribed to Cotugno (Liquor Cotunni) but the first clear description was provided some fifty years after Cotugno's report by Magendie (1825)

ANATOMICAL CONSIDERATIONS

The dural and the arachnoid membranes form the meningeal covering of the sac which contains the cerebrospinal fluid. The fluid circulates in the subarachnoid space which is for the most part narrow but is widened into spaces (cisterna) at several points. The principal dilatations are the cisterna magna below the cerebellum and above the medulla, the cisterna pontis on the ventral aspect of the pons and the cisterna basalis, which contains the circle of Willis. The subarachnoid space is lined with flattened epithelial cells and, at intervals, particularly in association with arachnoid villi, clumps of phagocytic cells called *meningocytes*, which stain like the cells of the reticulo-endothelial system, are found

SITE OF FORMATION

In 1853 Faivre reported the results of a histological study of the villous projections of the choroid plexuses into the ventricles of the brain. He found evidence of secretory activity in the cells covering these vascular structures. While this observation served to turn the attention of physiologists from the concept of Haller and Magendie, who believed that the fluid was formed by the leptomeninges, the evidence that the choroid plexuses are the principal structures concerned was not obtained until quite recently. While further histological and pharmacological studies gave additional support for this view, physiological experiments and observations on human cases have provided more conclusive results. Some of the more significant results will be cited: (1) When the aqueduct of Sylvius is occluded an internal hydrocephalus is produced (Dandy and Blackfan, Frazier and Peet), (2) during an operation on a human case a clear fluid was observed exuding from a choroid plexus (Cushing), (3) a sustained outflow of fluid, similar in volume to that obtained from the subarachnoid space was secured

from a catheter inserted into the aqueduct of Sylvius (Weed), (4) a unilateral internal hydrocephalus can be produced by obstructing one foramen of Monro (Dandy) but there is no excess accumulation of fluid when the choroid plexus is removed from the lateral ventricle before the obstruction is produced. These points leave no doubt of the importance of the plexuses in the elaboration of the fluid, but there is suggestive anatomical evidence that the perivascular spaces (Virchow-Robin), and the ependymal cells of the ventricles and the spinal canal may participate to some degree in this process. Each blood vessel entering the nervous system is surrounded by a channel lined for a distance, varying with the caliber of the vessel, with mesothelial cells. These perivascular spaces were at one time thought to be connected with the lymphatic system but there is no evidence for this. The spaces can be traced along the blood vessels until a direct communication with nerve cells is demonstrated. Thus large perivascular area provides a means of communication for fluid between the subarachnoid space and nervous tissue.

CIRCULATION

The fluid formed in the lateral ventricles passes through the foramen of Monro to join that produced in the third ventricle and thence through the aqueduct of Sylvius to the fourth ventricle. There is little doubt that the foramina in the roof of the fourth ventricle are true openings and not artefacts. The central one is the foramen of Magendie and the lateral openings bear the name of Luschka. Through these channels the fluid passes into the subarachnoid space and reaches the large cisternal dilatation (cisterna magna) which is situated at the medial cerebello-bulbar angle. From this cisterna the fluid passes slowly down the spinal canal within the arachnoid membrane and then, with some loss due to absorption and some gain due to formation in the cells lining the channel, it returns to the cerebral subarachnoid space. The circulation upward from the cisterna magna is somewhat more rapid, and the fluid bathes the base of the brain, the cerebral hemispheres, and indeed the whole central nervous system (see fig 71 1)

ABSORPTION

When a readily diffusible dye is introduced into the subarachnoid space its rapid appearance in the blood of the venous sinuses under certain conditions demonstrates a possible path of absorption of the fluid. When the dye is injected into the cisterna magna and the spinal canal is blocked, the absorption is not significantly lessened. This indicates that the fluid is largely absorbed from the cranial subarachnoid spaces. Key and Retzius believed that the fluid was absorbed through the Pacchionian granulations, but these are absent from the brains of infants and are now regarded as pathological enlargements of a few of the arach-

ral spaces into the lymphatic system was also demonstrated. These findings have been confirmed and it may be accepted that the main absorption of the cerebrospinal fluid is through the arachnoid villi into the great venous sinuses. The pathway postulated by Key and Retzius has therefore been established but the numerous microscopic arachnoid villi have been substituted for the Pacchionian granulations.

MECHANISM OF ABSORPTION Since the hydrostatic pressure in the subarachnoid space is always greater than that in the dural sinuses, filtration is apparently adequate to account for the flow of liquid into the venous blood stream. True solutions readily pass through the arachnoid villi, colloids more slowly, the rate depending upon the size of the molecule, particulate matter does not pass. No evidence of any secretory activity of the arachnoid villi has been obtained.

COMPOSITION AND MECHANISM OF FORMATION

In table 97 the amounts of the components of the cerebrospinal fluid are compared with those of



FIG 71.1 Lateral, horizontal view of the ventricular system. A, intraventricular foramina, B, foramen of Monro, C, anterior commissure, E, posterior commissure, F, pineal gland, G, aqueduct of Sylvius, H, fourth ventricle, I, (darker shadow) lateral recesses superimposed on the shadow of the fourth ventricle, J, superior posterior recess of the fourth ventricle, K, foramen of Magendie, L, tonsils of cerebellum, M, pons, N, medulla oblongata, O, anterior medullary velum, P, lingula of vermis of cerebellum, Q, posterior medullary velum, R, nodulus of vermis of cerebellum, S, lamina terminalis, T, choroid plexus and ependyma of roof of third ventricle, U, fornix, V, suprapineal recess (from Davidoff and Dyke)

noid villi. By long-continued slow injection of an isotonic solution of a mixture of potassium ferrocyanide and iron ammonium citrate, Weed was able to demonstrate that the particles of Prussian-blue formed when the tissue was subsequently fixed in acid medium, precipitated in the mesothelial cells of the tips of the arachnoid villi and within the dural sinuses into which these villi project. A relatively slow absorption by way of the perineu-

TABLE 97*

Comparisons of amounts of main constituents of blood plasma and cerebrospinal fluid

	BLOOD PLASMA	CEREBROSPINAL FLUID
	mg per 100 cc	mg per 100 cc
Protein	6300-8500	16-38
Amino acids	4.5-9	1.5-3
Creatinine	0.7-2.0	0.45-2.20
Uric acid	2.9-6.9	0.5-2.8
Cholesterol	100-150	absent
Urea	20-42	5-39
Sugar	70-120	45-80
Chloride (NaCl)	560-630	720-750
Inorganic phosphate	2-5	1.25-2.0
Bicarbonate (volumes per cent CO ₂)	40-60	40-60
Hydrogen ions (pH)	7.35-7.40	7.35-7.40
Sodium	325	325
Potassium	20	12-17
Magnesium	1-3	3-3.6
Calcium	9.0-11.5	4.0-7.0
Lactic acid	10-32	8-27

* Data largely that compiled by Flexner

blood plasma. The values given were obtained from analyses of lumbar fluid, that obtained from the ventricles may be slightly different. The water content of a unit weight of spinal fluid is somewhat greater than that of blood.

The question arises, how is the fluid formed? It will be apparent from the table that if certain constituents are investigated, diffusion from the blood plasma will be sufficient to account for the findings. However, before a decision can be reached as to whether the process is one of simple diffusion or of secretion, the concentrations of all the substances on both sides of the semipermeable membrane (the choroid plexus) must be determined. This was done by Flexner whose findings indicate that work must be done to form this fluid, i.e., that it is not merely a filtrate from blood plasma. Since the hydrostatic pressure of the capillary blood is believed to be greater than that of the spinal fluid, except under grossly abnormal conditions, this factor and the secretory power of the cells of the plexus are the two forces to be considered. Thermodynamic considerations indicate that the hydrostatic pressure difference would provide only a small fraction (about one-thirteenth) of the energy necessary. This leads to the conclusion that ultrafiltration will not account for the formation of the cerebrospinal fluid and that the cells of the choroid plexus perform work in this process.

THE AMOUNT AND PRESSURE

Accurate figures for the volume of the fluid in the various age groups are not available. The methods used to investigate the problem involve a change of conditions which invalidate the results secured. The total volume in healthy adults has been given as approximately 130 cc. The normal rate of formation has not been established. When artificial drainage is provided very large volumes, several liters per day, may drain away. The results of the urinary excretion of dye injected into the cerebrospinal space suggest that the volume of fluid is renewed every three or four hours but this rate of formation is probably too high. In recent experiments in which precautions were taken to interfere as little as possible with normal pressure conditions, Flexner and Winters found in experiments with adult cats, that approximately 12 cc. of fluid per day could be collected from a cannula fixed in the aqueduct. The pressure of the cerebrospinal fluid may be taken as 110 to 130 mm. of Ringer's for man in the recumbent position. Fluid drops at the rate of approximately one drop per second from the lumbar puncture needle. Pressure on the internal jugular vein or the rise of venous pressure produced by crying or coughing causes an increase in pressure of the

fluid presumably by increasing the size of the capillary bed in the brain. If the rise in venous pressure is long maintained there may also be retardation of the absorption of fluid.

EFFECT OF FLUID AND SALT INJECTIONS ON PRESSURE (WEED AND HUGHSON) When large volumes of isotonic solutions are injected intravenously there is a transient rise in venous and cerebrospinal pressure. Hypotonic solutions cause a prolonged rise in fluid pressure due presumably to the passage of fluid into the brain. There is a less marked and more transient rise of venous pressure.

Hypertonic solutions (30 per cent NaCl) administered intravenously, produce a fall in cerebrospinal pressure which may persist for long periods. Fluid is apparently attracted by the raised osmotic pressure of the blood from the brain substance, perivascular spaces, etc., into the blood stream. There is evidence also (Foley) that the direction of flow of the cerebrospinal fluid may be reversed. When the aqueduct of Sylvius is obstructed and intraventricular pressure is measured during the injection of the hypertonic solution, a fall in the pressure within the ventricles is noted. Using Weed's Prussian-blue, Foley was able to show that the dye passed from the ventricles into the capillaries of the choroid plexuses. When the aqueduct is open there is a current of fluid from the subarachnoid space into the ventricles. These phenomena are attributable to the raised osmotic pressure in the capillaries of the choroid plexuses.

Weed and Hughson's work was soon applied to the clinic. Intravenous injection of hypertonic saline has greatly facilitated brain operations by causing shrinkage of the brain and thus preventing extrusion of brain substance through trephine openings. It is obvious also that raised intracranial pressure resulting from various causes may be favorably affected, for a time at least, by the withdrawal of fluid from the brain. The raised osmotic pressure of the blood is usually produced by intravenous administration of saline but the oral route may be used (Cushing).

EFFECT OF CHANGE OF POSITION ON PRESSURE It is important to realize that the fluid in the subarachnoid space of the brain and spinal cord may be regarded as a single column (approximately 600 mm. long in the man of medium height). There are no structures to act as valves and restrictions of flow are not sufficient appreciably to interfere with this relationship. The change in pressure produced by a change of position can largely be predicted by hydrostatic considerations when the

blood pressure remains constant, but the elasticity of the dural sac varies in different individuals. The pressures in the lumbar and occipital regions are identical when the subject is horizontal. The pressure in the lumbar region in man is approximately 200 mm higher in the sitting than in the prone position. The pressure in the human cisterna magna in the erect posture is probably below atmospheric. Experimental data indicate that some point in mid-thoracic region has a pressure coinciding with that of the atmosphere. The further exploration of these questions gives promise of important advances in our knowledge of the hydrostatics of the cerebrospinal fluid (Weed). The hypothesis formulated by Monro and supported by Kellie (the Monro-Kellie doctrine) postulated constancy of cerebral volume, i.e., a reciprocal relationship between the volumes of blood and spinal fluid. The third substance within the dural sac, the brain cells, are usually assumed to remain of constant volume. This hypothesis has been restated by Weed and now includes an "elastic" factor which takes into consideration a component depending on the ease of vascular adjustments as well as the distensibility and collapsibility of the meningeal sac.

FUNCTION OF THE CEREBROSPINAL FLUID

The fluid within the elastic meningeal sac serves as a protective covering for the nerve cells. By change in its volume compensation for change in the amount of blood is effected and the contents of the cranium thus tend to remain of constant volume. There is probably considerable exchange of metabolic materials between the nerve cells and the fluid.

VENTRICULOGrams AND ENCEPHALOGrams

The location of a bubble of air injected into the cerebral ventricles (ventriculography) or through

a lumbar puncture (encephalography) can be visualized by the X-rays. The air should rise to the top of the column of fluid and follow change in position of the subject. Failure to do this suggests abnormality of the communicating channels.

Lipiodol, an iodized vegetable oil, opaque to X-rays, is sometimes injected into the cisterna magna. Normally it falls to the lower region of the spinal canal but narrowing of the canal by tumor or adhesions may prevent its passage and the level of the obstruction can in this way be determined from X-ray pictures.

HYDROCEPHALUS

A consideration of the formation, circulation and absorption of the spinal fluid leads to the assumption that hydrocephalus might be produced by (1) Increased rate of formation of fluid. A very definite increase can be produced by the administration of hypotonic solutions (Weed) but the decreased tonicity of the blood cannot be maintained long enough to permit the development of hydrocephalus. Clinically hypertrophy of the choroid plexuses might produce this condition and a suggestive case has been reported (Davis). (2) Obstruction to the passage of the fluid. An obstruction by tumor or inflammation in the right foramen of Monro, for example, would produce a right internal hydrocephalus. Blockage of the aqueduct of Sylvius would produce a bilateral internal hydrocephalus as would an obstruction in the fourth ventricle or at the foramina of Magendie and Luschka. (3) Interference with the absorption of fluid by way of the arachnoid villi. Blockage of many of these villi or interference with their function by other means may lead to an external hydrocephalus. Increased intracerebral venous pressure may produce a temporary decrease in absorption of fluid.

CHAPTER 12

THE AUTONOMIC NERVOUS SYSTEM

(Synonyms— involuntary nervous system,
vegetative nervous system)

The autonomic nervous system has been touched upon in many of its aspects in other sections of this book. There remains to be given an account of the structural plan of this system as a whole, a general summary of its functions and of the structures which it innervates. From anatomical, physiological and pharmacological viewpoints the autonomic system falls naturally into two main divisions—the *sympathetic* or *thoracolumbar outflow* and the *parasympathetic* or *craniosacral outflow* (see fig 72 1)

THE SYMPATHETIC DIVISION

The cells of origin of the sympathetic division are situated in the lateral horns of the spinal cord (intermediolateral cell column) from the 8th cervical or 1st thoracic to the 2nd or 3rd lumbar segments. The axons of these cells leave the cord by the corresponding anterior nerve roots and synapse with nerve cells in one or another of the outlying ganglia. The fibers arising from the spinal cells are medullated and are called *preganglionic*, those arising from cells of the ganglia are non medullated, and are called *postganglionic*. Evidence for the existence of a higher center in the posterior region of the hypothalamus has been considered in chapter 67, and in the cerebral cortex in chapter 68.

The ganglion cells are all motor or secretory in function (p 1095). They have no afferent connections.

The ganglia are arranged in three systems or groups (A) *vertebral* (or *central*), (B) *prevertebral* (or *collateral*) and (C) *terminal* (or *peripheral*).

A. THE VERTEBRAL (OR CENTRAL) GANGLIA AND THE GANGLIATED CORD

The vertebral group lies in close relation to the vertebral bodies and consists, on each side, of a series of some 22 ganglia connected together by intervening fiber tracts to form a nodular cord extending from the base of the skull to the front of the coccyx. This is known as the *sympathetic chain* or the *gangliated cord of the sympathetic*. It will be described in sections

The cervical part of the sympathetic chain

The cervical part of the sympathetic chain possesses three ganglia—the *superior*, *middle* and *inferior cervical ganglia*. They are relatively large and are believed to result from the fusion of two or more smaller ganglia.

THE SUPERIOR CERVICAL GANGLION, situated below the base of the skull, is the largest of the three. It receives preganglionic fibers from the upper thoracic segments of the cord, and its cells supply fibers (postganglionic) to the vessels, glands and cutaneous muscle of the head. It is probably formed by the fusion of the uppermost three or four cervical ganglia. Its branches are

(1) THE INTERNAL CAROTID NERVE. This nerve, composed of postganglionic fibers, arises from the upper pole of the superior cervical ganglion and, passing into the cranium with the artery of the same name, forms the internal carotid and cavernous plexuses.

(i) The *internal carotid plexus*, situated on the lateral side of the internal carotid artery, sends branches to the following

(a) The abducent nerve.

(b) The tympanic branch of the glossopharyngeal nerve.

(c) The sphenopalatine ganglion. These fibers pass by way of the deep petrosal nerve and the nerve to the pterygoid canal which is formed by the union of the former nerve with the great superficial petrosal nerve. Orbital branches of the sphenopalatine ganglion convey sympathetic fibers to the lacrimal gland, the soft palate, nasopharynx and pharynx receive fibers through the palatine and pharyngeal branches of the ganglion.

(d) The semilunar (trigeminal) ganglion.

The sympathetic fibers pass through the sphenopalatine and semilunar ganglia without interruption.

(ii) The *cavernous plexus*, situated on the inner side of the internal carotid artery as it lies in the cavernous sinus, sends branches to the following

(a) The oculomotor, trochlear and abducent nerves and the nasociliary branch of the ophthalmic division of the trigeminal nerve. Through the long ciliary nerves (twigs of the nasociliary nerve), sympathetic fibers are conveyed to the dilator pupillae (see also p 1174).

(b) The ciliary ganglion, through which the sympathetic fibers pass without interruption into the short ciliary nerves. These fibers provide an additional pathway for sympathetic impulses to the dilator of the pupil.

(c) The pituitary body (vasomotor)

Through the communicating branches of the cavernous plexus the vessels of the eyeball and nasal mucosa are supplied with constrictor fibers, and the skin of the nose with vasoconstrictor, motor (smooth muscle) and secretory (sweat) fibers. The terminal filaments

(4) Filaments to the CAROTID SINUS and CAROTID Body

(5) Fibers to LARYNGEAL and PHARYNGEAL PLEXUSES

(6) The SUPERIOR CARDIAC NERVE to the cardiac plexuses

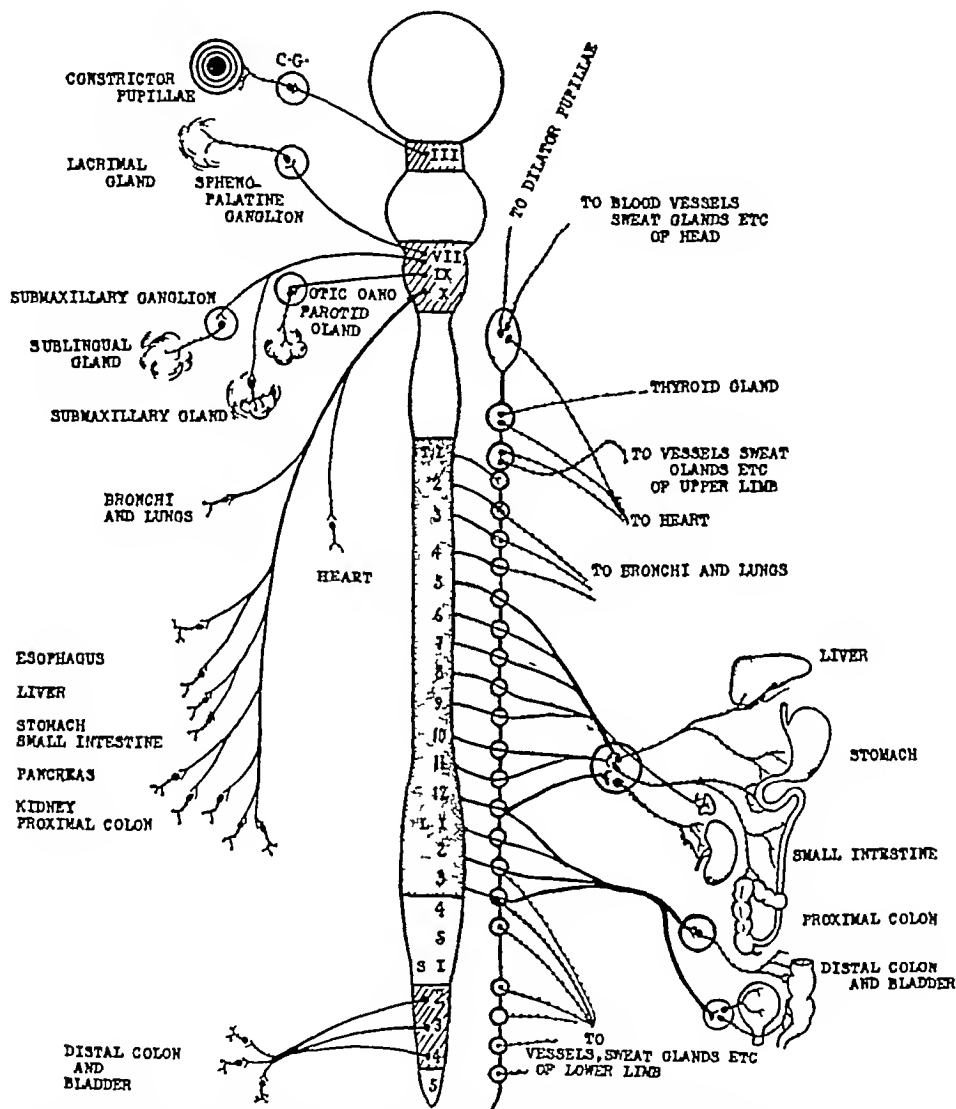


FIG 72.1 Showing plan of autonomic nervous system from a functional viewpoint. C G, ciliary ganglion. The celiac, inferior mesenteric and hypogastric ganglia are represented, in this order from above downwards, by the circles in the lower right portion of the diagram

of the internal carotid and cavernous plexuses are continued as delicate networks over the anterior and middle cerebral arteries to the minute vessels of the pia mater, and along the ophthalmic artery to the structures of the orbit.

(2) Branches to and distributed with the UPPER FOUR CERVICAL NERVES

(3) Twigs to the GANGLION JUGULAR AND GANGLION NODOSUM OF THE VAGUS, to the PETROUS GANGLION OF THE GLOSSOPHARYNGEAL NERVE and to the HYPOGLOSSAL NERVE

(7) Branches which ramify in a plexiform manner upon the external carotid artery—the EXTERNAL CAROTID PLEXUS. This plexus is continued over the branches of the external carotid and supplies fibers to the vessels, sweat glands and cutaneous muscles of the face, and to the thyroid gland. Filaments from the plexus investing the facial artery pass to the submaxillary ganglion. The plexus on the middle meningeal artery sends fibers to the otic ganglion, these pass without interruption into the auriculotemporal nerve through which they reach the parotid gland.

THE MIDDLE CERVICAL GANGLION is formed presumably by the coalescence of the fifth and sixth cervical ganglia. Its branches are as follows:

(1) Branches to the fifth and sixth cervical nerves and thence to the blood vessels, sweat glands and cutaneous muscle within the area of distribution of these nerves.

(2) The middle cardiac nerve to cardiac plexuses.

(3) Branches which extend along the inferior thyroid artery to the thyroid and parathyroid glands.

THE INFERIOR CERVICAL GANGLION probably represents the union of the seventh and eighth cervical ganglia, in the dog and cat, and in most human subjects, it is often fused with the first thoracic ganglion, and occasionally with the second as well, to form the so-called *stellate ganglion*. It gives off the following branches:

(1) Branches to the seventh and eighth cervical and the first thoracic nerves, sometimes also to the sixth cervical and the second thoracic.

(2) The inferior cardiac nerve to cardiac plexuses.

(3) Branches which form plexuses upon the subclavian artery and its branches. Sympathetic fibers are thus carried into the cranial cavity along the vertebral artery, and over the axillary and commencement of the brachial (see also ch. 27).

The thoracic, lumbar and sacral ganglia

The *thoracic ganglia* are 10 or 12 in number on each side. They are evenly spaced, one to each spinal segment. As mentioned above the first thoracic and inferior cervical ganglia are commonly partially or completely fused to form an irregularly shaped mass, called the *stellate ganglion*. There are usually 4 lumbar and 4 or 5 sacral ganglia.¹ The sacral portions of the two sympathetic trunks converge below and fuse upon the anterior surface of the coccyx to form a terminal swelling—the *coccygeal ganglion* or *ganglion impar*.

¹ Small accessory ganglia, called *intermediate ganglia*, are found outside the sympathetic chain proper, attached to the rami communicantes close to the spinal nerve roots in the cervico-thoracic and lumbar regions. From these ganglia sympathetic postganglionic fibers proceed for distribution by the brachial and lumbar plexuses to the limb vessels. Preganglionic fibers leaving the cord by the anterior roots and entering the intermediate ganglia without passing through the sympathetic chain offer a possible alternative pathway for sympathetic impulses to the limbs. Such fibers would remain intact though the sympathetic chain were excised, their presence would, thus, offer an explanation in some instances for the failure of sympatheticectomy to completely denervate the vessels of the limbs.

B THE PREVERTEBRAL (OR COLLATERAL) GANGLIA

These lie in the thorax, abdomen and pelvis in relation to the aorta and its branches. The larger of the prevertebral ganglia are (a) the *celiac (solar or semilunar) ganglion*, lying in relation to the origin of the celiac artery, (b) the *superior mesenteric ganglion*, situated below the origin of the superior mesenteric artery, and (c) the *inferior mesenteric ganglion*, which bears a corresponding relation to the inferior mesenteric artery, this ganglion is rarely present in man (see also p. 1094).

C THE TERMINAL GANGLIA

These consist of small collections of ganglion cells situated in close relation to the innervated organs, especially those of the pelvis, e.g., the bladder and rectum.

THE OUTFLOW OF SYMPATHETIC FIBERS FROM THE CENTRAL NERVOUS SYSTEM

It has already been stated that the cells giving rise to the sympathetic fibers (p. 1088) are situated in the thoracic and upper lumbar segments of the cord. It is from this limited region (8th C or 1st T to 2nd or 3rd L inclusive) that the sympathetic (preganglionic) fibers emerge. *This region constitutes the only outlet for sympathetic impulses.* So the term *thoracolumbar outflow* simply means the sympathetic division of the autonomic nervous system. The fibers emerge from the cord through the anterior root of the spinal segment in which their cell bodies are placed. In a cross section of the anterior root they appear as fine medullated fibers (2.5 μ or less in diameter) intermingled with the larger, medullated, somatic (motor) fibers. They separate almost immediately, however, from the voluntary motor fibers of the anterior root and enter the corresponding ganglion of the sympathetic chain. Thus, the spinal nerves from the 8th cervical or 1st thoracic to the 2nd or 3rd lumbar, but not others, are connected each to a vertebral ganglion by a delicate white strand composed of preganglionic fibers and known as the *white ramus communicans* (plural, *rami communicantes*, fig. 72.2 and fig. 27.1, p. 274). A preganglionic fiber after entering the ganglion may pursue one of three courses: (a) form synapses with cells in the ganglion which it first enters, (b) pass up or down the sympathetic trunk for some distance to terminate in a ganglion at a level higher or lower than that of the segment from which it originated. It

may give off collateral branches to ganglion cells along its course. In any event, the preganglionic fibers issuing from a given segment connect with several ganglia (from five to nine). Furthermore, each preganglionic fiber may form a large number of synapses within a given ganglion. Ranson and Billingsley found that in the case of the superior

of the ganglion cells which are therefore called *postganglionic*. The gray rami in the thoracic and upper lumbar regions join the spinal nerves close to the points at which the white rami arise. Their constituent fibers are continued in the peripheral nerves for the supply of the blood vessels, sweat glands and smooth muscle of the skin. Whereas,

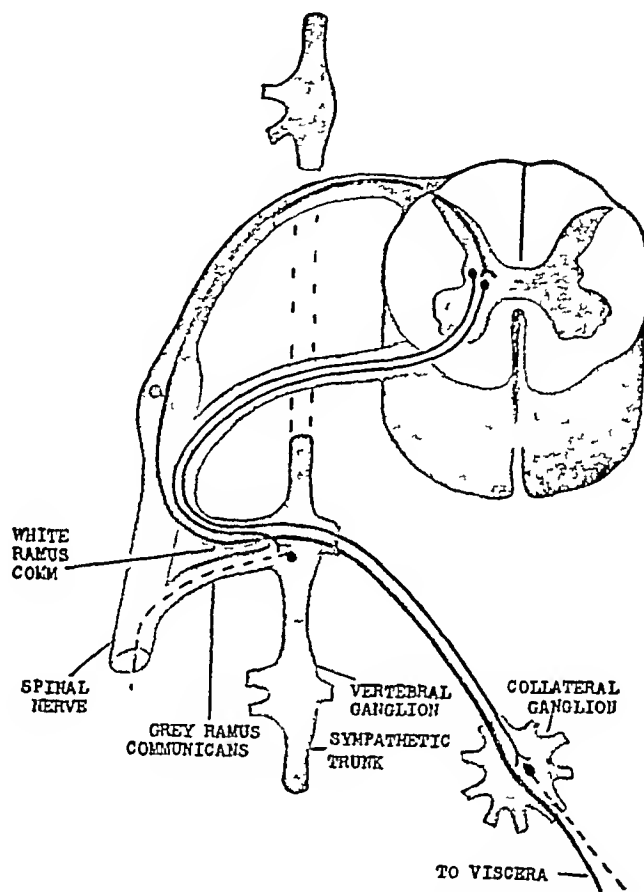


FIG 72.2 Diagram showing the connections of sympathetic fibers. Efferent fibers in black, preganglionic, solid lines, postganglionic, interrupted lines. Afferent visceral fiber in red.

cervical ganglion each fiber communicated with some thirty-two ganglion cells, this accounts for the diffuse nature of the sympathetic discharge. (c) Traverse the gangliated cord without interruption to find a cell station in either a prevertebral or a terminal ganglion.

PLAN OF DISTRIBUTION OF SYMPATHETIC FIBERS TO THE PERIPHERY

To the limbs and trunk

The ganglia of the sympathetic chain are connected with the spinal nerves supplying the limbs and trunk by delicate filaments called *gray rami communicantes*. They are composed of the axons

as already mentioned, only a limited number of spinal nerves possess white rami, *every spinal nerve receives a gray ramus*. It therefore follows that sympathetic impulses going to parts supplied by upper cervical, lower lumbar or sacral somatic nerves, must travel considerable distances up or down the sympathetic trunk before reaching an outlet through a gray ramus. The preganglionic fibers to the upper limb arise from the 2nd to the 7th spinal segments, most of these fibers are contained in the white rami of the 2nd and third thoracic nerves. The postganglionic fibers to the upper limb arise from the first thoracic, and the inferior and middle cervical ganglia (see also ch

27), those for the lower limb from the lumbar and sacral ganglia. Excision of the inferior cervical and first thoracic ganglia, or the stellate ganglion when present interrupts the sympathetic pathways to the upper limb as well as those to the head and neck and most of those to the heart. This operation may fail, however, completely to remove the sympathetic supply to the arm since a gray ramus (Kuntz's nerve) frequently passes from the second thoracic ganglion via the first thoracic nerve to join the brachial plexus. In man, virtually complete sympathetic denervation of the upper limb can be effected by sectioning the rami communicantes of the 2nd and 3rd thoracic nerves, and dividing the sympathetic cord below the 3rd ganglion.

To the head and neck

Sympathetic impulses to the structures of the neck—face, scalp and intracranial cavity—are conveyed from the cord in the white rami of the upper two thoracic nerves. They ascend to connect with cells in the middle and superior cervical ganglia. From the latter postganglionic fibers are distributed through the internal carotid, cavernous and external carotid plexuses as described on page 273. The sympathetic as well as supplying the blood vessels, sweat glands and pilomotor muscles of the head and neck also innervates the salivary glands, the dilator pupillae, Mueller's muscle and the smooth muscle component of the levator palpebrae superioris. Sympathetic fibers also descend the infundibular stalk to the pituitary gland. The spinal center for the dilator pupillae is situated (in man) in the 8th cervical or 1st thoracic segment. The preganglionic fibers are found in the white rami of the first thoracic nerve, and are distributed, as already stated through the internal carotid and ciliary nerves.

To the thoracic viscera

The sympathetic postganglionic fibers join with the branches from the vagus to form the *cardiac, pulmonary and esophageal plexuses*.

The *cardiac plexus* lies in relation to the origins of the aorta and pulmonary artery. It consists of a superficial and a deep portion, and is formed by the interlacement of fibers from the cardiac branches of the vagus (parasympathetic) and sympathetic nerves. The vagus fibers are preganglionic. They terminate around ganglion cells in the walls of the heart (p. 241). The sympathetic fibers derived from the superior, middle and in-

ferior cardiac nerves are postganglionic, their cell stations lying in the corresponding cervical ganglia (fig. 25.2, p. 244). The preganglionic fibers arise from the upper four or five thoracic segments of the cord.

The *pulmonary plexuses*, anterior and posterior, lie in relation to the corresponding aspects of the root of the lung. They are formed from postganglionic fibers of the sympathetic (T. 2, 3 and 4) and preganglionic fibers of the vagus. The latter connect with ganglion cells in the walls of the bronchi. Herein is situated an intrinsic nervous plexus consisting of these ganglion cells and medullated and non-medullated fibers.

The *esophageal plexus* embraces the lower half of the esophagus. Vagal and sympathetic fibers (from the upper thoracic ganglia and from the thoracic portion of the great splanchnic nerve) enter into its formation. The vagal fibers end around ganglion cells of the intrinsic plexus of Auerbach in the esophageal wall.

To the abdominal and pelvic viscera

The *greater, lesser and least splanchnic nerves*. These are composed of preganglionic fibers, and may be looked upon as elongated white rami. They connect with cells in the prevertebral (collateral) ganglia. The postganglionic fibers after emerging from the latter join the neighboring plexuses. The *greater splanchnic nerve* arises from the cord from as high as the 4th or 5th thoracic segment, and as low as the 9th or 10th. Its fibers end in the upper part of the celiac ganglion, from here postganglionic fibers are continued into the celiac plexus. The *lesser splanchnic* and the *least (or lowest) splanchnic nerves* are much smaller. The former arises from the 9th and 10th or the 10th and 11th thoracic segments and its fibers, after passing without interruption through the vertebral ganglia at these levels, end in the lower portion of the celiac (or aorticorenal) ganglion. The *least splanchnic nerve* arises from the last one or two thoracic segments and first lumbar segment, it joins the renal plexus (fig. 72.3). Postganglionic fibers arising from small ganglia within the plexus are distributed to the kidney and ureter.

The *lumbar splanchnic nerves* are three or four strands which arise from the second and third lumbar segments. Their fibers pass through the lumbar portion of the sympathetic chain and enter the inferior mesenteric ganglion, here some are relayed, others are continued without interruption and find their cell stations in peripheral ganglia.

THE PLEXUSES OF THE ABDOMEN AND PELVIS
The sympathetic fibers form rich plexuses in relation to the aorta and its branches from which filaments pass to the abdominal and pelvic viscera. Parasympathetic fibers also enter into the constitution of these plexuses.

plexuses—the *hepatic, gastric, splenic, renal* and *adrenal plexuses*—which invest the corresponding arteries and their branches.

The *superior mesenteric plexus* is continuous above with the celiac plexus. It surrounds the superior mesenteric artery, along the branches of which it is prolonged as subsidiary plexuses. The plexus is composed of post-

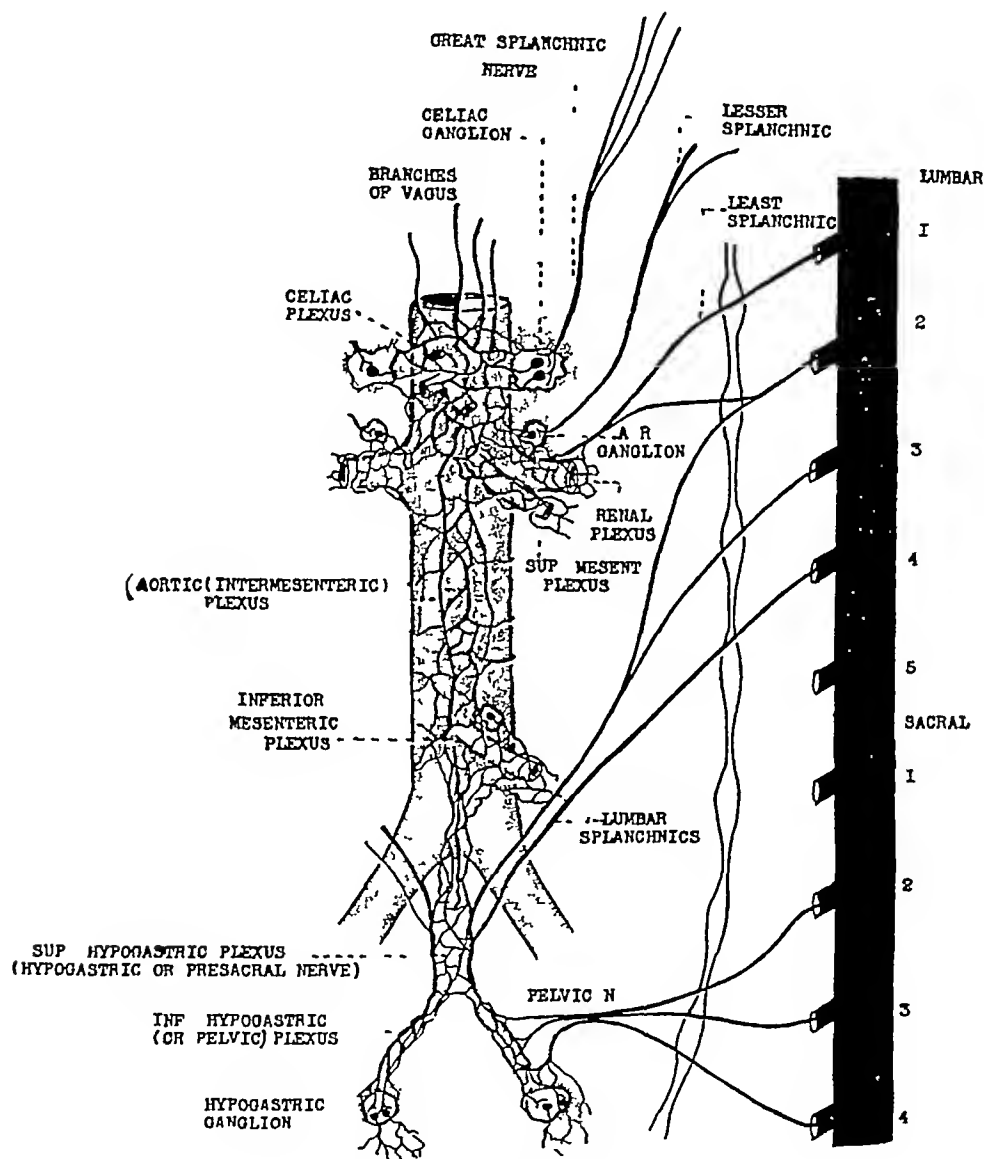


FIG 72.3 Diagram of the nerve plexuses of the abdomen and pelvis. A R., aortico-renal

The *celiac* (or *solar*) *plexus* lies upon the abdominal aorta at the origin of the celiac artery. The *celiac ganglia*, right and left, lie embedded within the plexus, which is made up of fibers arising in the ganglion (i.e., postganglionic fibers of the greater and lesser splanchnic nerves) together with preganglionic fibers of the vagus. The lower part of the celiac ganglion is often detached and is then referred to as the *aortico-renal ganglion*. The plexus invests the celiac artery throughout its course and gives rise to several subordinate

ganglionic fibers which arise in an aggregation of nerve cells—the *superior mesenteric ganglion*—lying within it, of preganglionic fibers derived from the lumbar segments of the cord, and from the celiac plexus. It and the other plexuses to be described also contain parasympathetic fibers. The superior mesenteric plexus supplies the pancreas, and the small intestine and the large intestines as far as the commencement of the descending colon.

The *aortic* or *intermesenteric plexus* lies upon the aorta

between the origins of the superior and inferior mesenteric arteries. It receives fibers from the celiac plexus and from the upper lumbar ganglia. The aortic plexus gives rise secondarily to the *spermatic* and *ovarian* plexuses which supply the testes and ovary with autonomic fibers. It is connected below with the *inferior mesenteric plexus* which invests the artery of the same name. The *inferior mesenteric ganglion*, a collection of ganglion cells lying within the latter plexus, receives the lumbar splanchnics, this ganglion is not present, as a rule, in the human subject (Learmonth), but a number of smaller ganglia are found scattered through the plexus. The plexus is formed of fibers derived from the aortic plexus and from the inferior mesenteric ganglion when this is present. From it secondary plexuses arise which invest the branches of the inferior mesenteric artery and carry sympathetic impulses to the descending colon, iliac colon, pelvic colon and rectum.

The *superior hypogastric plexus* is the downward extension of the aortic plexus. It lies over the lower end of the aorta, and in the angle formed by the aortic bifurcation. Though rarely condensed into a single bundle it is sometimes referred to as the *hypogastric nerve* or the *presacral nerve*. It transmits inhibitory impulses to the pelvic colon and via the pelvic plexuses to the rectum, bladder and other pelvic viscera. It divides below into the right and left *pelvic* or *inferior hypogastric plexuses*. These lie one on either side of the rectum and are composed of medullated and non-medullated fibers among which are scattered numerous small ganglia which are sometimes referred to collectively as the *hypogastric ganglion*. Parasympathetic fibers enter the inferior hypogastric plexuses through the pelvic nerve (sacral outflow, p. 1095). The sympathetic fibers contributing to the pelvic plexuses have their ultimate source in the lumbar segments of the cord. They reach the plexuses via the hypogastric plexuses as well as more directly from the sacral part of the sympathetic chain. Through subsidiary plexuses—*hemorrhoidal*, *vesical*, *uterine*, *vaginal* and *prostatic*—fibers (sympathetic and parasympathetic) are conveyed from the pelvic plexuses to the pelvic viscera.

THE PARASYMPATHETIC OR CRANIOSACRAL DIVISION OF THE AUTONOMIC NERVOUS SYSTEM

The cells giving rise to parasympathetic fibers are situated at three different levels of the central nervous system—the *mid brain*, the *medulla* and the *sacral region of the spinal cord*. The axons of these cells leave the central nervous system to connect with ganglion cells lying within or in close relation to the innervated organ (see fig. 65.9). As in the case of the sympathetic division the axons of the central cells are called *preganglionic*, those of the ganglion cells, *postganglionic*. The former are medullated, the latter non-medullated.

The three levels from which parasympathetic fibers emerge will be referred to as the *tectal* (or *mid brain*), the *bulbar* and the *sacral* outflows, respectively.

A THE TECTAL OR MID BRAIN OUTFLOW

The group of cells comprising the Langer-Westphal nucleus of the oculomotor nerve (p. 1181), in the floor of the cerebral aqueduct, are believed to give rise to the tectal fibers.

The autonomic fibers are conveyed in the third nerve as far as the ciliary ganglion where they find their cell stations. Postganglionic fibers emerge from the ganglion in the *short ciliary nerves*, and terminate in the sphincter pupillae and the ciliary muscle (p. 1174).

B THE BULBAR OUTFLOW

These fibers leave the brain in the *facial*, *glossopharyngeal* and *vagus nerves*.

(1) The parasympathetic fibers (secretory and vasodilator) entering the *facial nerve* arise from the *superior salivatory (salivary) nucleus*, which lies dorsal and lateral to the lower end of the motor nucleus of the facial.² These fibers emerge from the brain in the sensory root of the facial nerve (*nerve intermedius*, ch. 66) and travel with the latter to the facial canal of the temporal bone. Here they leave the facial, (i) in its *chorda tympani* branch which later joins the lingual to be conveyed to the floor of the mouth. At this point some of the chorda fibers (secretory and vasodilator in function) separate from the lingual again to enter the *submaxillary (submandibular) ganglion* from where they are relayed, the postganglionic fibers pass to the submaxillary and submaxillary glands, and the mucous membrane of the mouth. (ii) In the great superficial petrosal nerve and nerve of the pterygoid canal (Vidian nerve) to the sphenopalatine ganglion. From here postganglionic fibers pass via orbital branches of the ganglion to the lacrimal gland, and to the mucous membrane of the soft palate, nasopharynx and pharynx via the palatine nerves. Vasodilator fibers also leave the facial by the great superficial petrosal nerve, and entering the cranium are conveyed along the middle meningeal artery and its branches.

(2) The parasympathetic fibers (secretory and vasodilator) of the *glossopharyngeal nerve* arise from cells of the *inferior salivatory (salivary) nucleus*. This nucleus lies in the uppermost part of the medulla immediately below the superior salivatory nucleus, and lateral to the motor nucleus of the glossopharyngeal nerve. The autonomic fibers leave the brain with the latter nerve but separate from it again in its tympanic branch (Jacobson's nerve) which joins a twig from the genicu-

² This nucleus is actually in the lowest part of the pons.

late ganglion (g of facial nerve) and filaments from the internal carotid plexus to form the *tympanic plexus*. From this plexus emerges the *small superficial petrosal nerve* through which parasympathetic fibers are continued to the otic ganglion. From this ganglion postganglionic fibers are conveyed to the parotid gland via the auriculotemporal nerve (p 492). The tympanic plexus itself sends filaments to the mucous membrane of the tympanic cavity, the mastoid air-cells, auditory (Eustachian) tube and the internal ear.

(3) The *vagus nerve* contains the greater proportion of the fibers of the bulbar outflow. They arise from the *dorsal nucleus* of the vagus and are distributed through the latter's numerous branches to the thoracic and abdominal viscera (ch 66). Unlike those in the other two cranial nerves, the preganglionic fibers of the vagus connect with ganglion cells situated within the innervated organs. Thus the vagal fibers to the heart connect with ganglion cells in the cardiac walls; those to the bronchi with the nerve cells of the intrinsic plexus in the bronchial walls; those to the esophagus, stomach and intestine form synapses with the ganglion cells of the myenteric plexus of Auerbach and the submucous plexus of Meissner. The preganglionic fibers are therefore quite long, the postganglionic very short. The cells of origin of most of the cardiac fibers of the vagus form a discrete group (cardio-inhibitory center) lying alongside the dorsal nucleus of the vagus.

C THE SACRAL OUTFLOW

The cells of origin lie in the anterior horns of the 2nd, 3rd and 4th and sometimes the 1st sacral segments of the cord. The preganglionic fibers emerge in the anterior roots of the corresponding sacral nerves. The fibers leave the spinal nerves again and, proceeding peripherally as the *pelvic nerve* (or *nervus erigens*), on each side, enter into the formation of the pelvic plexus. The fibers terminate around ganglion cells lying in close relation to the pelvic organs. They carry motor impulses to the walls of the descending colon, rectum and bladder, inhibitory impulses to the internal anal and vesical sphincters and to the uterus, and dilator impulses to the blood vessels of the bladder, rectum and genitalia.

THE AFFERENT VISCERAL NERVES

Impulses are transmitted from the viscera by afferent fibers which pass through the various plexuses and reach the central nervous system via the vagus, pelvic, splanchnics and other autonomic nerves.

The *afferent fibers of the sympathetic division* are the peripheral processes of ganglion cells in the posterior spinal nerve roots from the 1st thoracic to the 3rd lumbar segments. *None arise from*

sympathetic ganglia. They reach the sympathetic trunk via the white rami communicantes and are distributed to the viscera along with the corresponding efferent fibers (fig 72.2). Though some of these afferent fibers are non-medullated the majority are medullated, and of larger size than the efferent fibers. The ganglion cells of the posterior roots which give origin to the afferent fibers of the sympathetic (or of the pelvic nerve) have not been shown definitely to differ from those giving rise to the ordinary somatic sensory fibers. For this reason the sympathetic system proper is sometimes regarded, though perhaps irrationally, as consisting solely of efferent neurons. The afferent fibers pass to their destinations without interruption in any of the sympathetic ganglia, passing directly to the viscera in the splanchnics or the other visceral nerves. A certain proportion also enter the spinal nerves for distribution to the limbs via the gray rami.

The *Afferent fibers of the vagus* are the peripheral processes of cells in the ganglion jugulare and the ganglion nodosum. The central processes of these neurons terminate in the dorsal nucleus of the vagus (ch 66). Therein connections are made with efferent parasympathetic neurons to complete the reflex arc. The *afferent fibers of the pelvic nerve* arise from cells in the posterior root ganglia of the 2nd, 3rd and 4th sacral nerves. They pass peripherally with the efferent autonomic fibers.

The *visceral reflex arc*, as pointed out by Gaskell, is formed upon a plan similar to that upon which somatic reflexes are based. The afferent fiber in the latter instance is connected to the anterior horn cell through the intermediary of an intraspinal (internuncial) neuron or a series of such neurons. These are spoken of as connector neurons. In the case of visceral reflexes, the afferent fiber makes contact with a cell in the lateral horn of gray matter. The axon of this cell—the preganglionic fiber—which connects with a ganglion cell of the sympathetic system, corresponds to the connector fiber of the somatic reflex arc. In development, however, this neuron has migrated from the central nervous system, being only in part intraspinal. The neuron with which it communicates, i.e., the ganglion cell and postganglionic fiber, though it has been carried entirely beyond the bounds of the central nervous system, corresponds to the motor neuron of the somatic reflex arc.

A certain degree of independent reflex activity can be carried out through the intrinsic plexuses, e.g., of the intestine, when these are separated

from the central nervous system by division of the main autonomic nerves. It is also true that some independent activity can be carried out through axon reflexes or possibly through some of the more peripherally placed ganglion cells. Other parts of the autonomic system, however, cannot function apart from the central nervous system. The larger ganglia of the parasympathetic or sympathetic, for instance, do not serve as reflex centers. It is clear from the description of the origin and course of the visceral afferent nerves given above that no anatomical basis for such action exists.

THE FUNCTIONS OF THE AUTONOMIC NERVOUS SYSTEM

The autonomic nervous system governs the activities of cardiac and smooth muscle, of the digestive glands and sweat glands of the adrenal medulla and possibly of certain endocrine organs. It is concerned with those processes which normally are beyond voluntary control and are for the most part beneath consciousness. The term autonomic as applied to this system is not altogether suitable since, as we have seen, it is under the control of centers within the central nervous system, and cannot function as an independent unit.

Through its various activities the autonomic system exercises the important function of maintaining the constancy of the fluid environment of the body's cells, it serves to combat forces, acting either from within or without, which tend to cause variations in this environment. Regulation of the composition of the body fluids, of their temperature, quantity and distribution, is effected through the actions of the autonomic nerves upon circulatory, respiratory, excretory and glandular organs. For example, of glandular structures under autonomic influence, the liver, pancreas and adrenal medulla are of especial importance in the regulation of blood sugar, the sweat glands aid in the control of body temperature. The neural lobe of the pituitary, in most of its functions at least, is also under autonomic nervous control. On the contrary, the thyroid gland, adrenal cortex, adenohypophysis and probably also the parathyroid glands, are governed by hormones liberated by the pituitary (ch 57), though it is not improbable that they may be influenced also by impulses received through autonomic paths. The stability of the internal environment (the *milieu interne* of Claude Bernard) which is so characteristic of the healthy body, is spoken of by Cannon as *homeo-*

stasis. According to Cannon, the essential and particular function of the autonomic system is to bring about the internal adjustments upon which this constant state depends. He therefore refers to the autonomic nerves as the *interofective* system. He speaks of the voluntary system (i.e., the central nervous system and the somatic nerves) as the *exteroffective* system, since through its exteroceptors and effectors a direct relationship is established with the external environment.

The great majority of the effector organs of the autonomic system are innervated by both sympathetic and parasympathetic divisions (see table 98), and the effects exerted by the two types of fiber going to a given organ are antagonistic. Thus the heart's action is inhibited by the vagus but augmented by the sympathetic. In the intestine the effects of the two nerves are reversed, the parasympathetic (vagus nerve to the small bowel and upper half of the colon, and the pelvic nerve to the lower half of the latter) are augmentor, the sympathetic is inhibitory. The removal of the effects of one set of fibers, as by section, results, as a rule, in the effects of the other set becoming more prominent. This fact indicates that each type of fiber exerts a constant or tonic action and suggests that the two effects are delicately balanced one against the other. Thus, section of the vagus nerves causes an increase in the cardiac rate, and section of the parasympathetic or of the sympathetic fibers to the iris causes, respectively, dilatation or constriction of the pupil. See pupillary reactions, (p 1174).

Taken as a whole the actions of the sympathetic division and its ally the medulla of the adrenal gland (sympathoadrenal system, ch 59) are directed towards strengthening an animal's defences against the various dangers which beset it, e.g., extremes of temperature, deprivation of water or the attacks of its enemies. It has been shown by Cannon, however, that the sympathetic system is not indispensable, both gangliated cords may be completely removed yet the animal remains in good health provided it is kept in the sheltered environment of the laboratory. Sympathectomized cats if kept warm and carefully tended will live indefinitely. Kittens deprived of their sympathetic trunks grow normally, female cats become pregnant and give birth to young, though the mammary glands do not function normally and the maternal instinct is lacking. Sympathectomized animals are, however, incapable of arduous work, sugar is not mobilized from the liver on demand,

TABLE 98

ORGAN	PARASYMPATHETIC EFFECTS*	ORIGIN OF SYMPATHETIC POSTGANGLIONIC FIBERS	SYMPATHETIC EFFECTS
<i>Heart</i> (p 241, 243)	Inhibition	Superior middle and inferior cervical ganglia	Acceleration
<i>Vessels</i>			
Cutaneous (p 272-279 and 312)	—	Various vertebral ganglia	Constriction
Muscular (p 275)	—	Various vertebral ganglia	Dilatation, Const
Coronary (p 326)	Constriction/	Cervical ganglia	Dilatation
Salivary glands (p 489)	Dilatation	Superior cervical ganglion	Constriction
Buccal mucosa	—	Superior cervical ganglion	Dilatation
Pulmonary (p 333)	Dilatation and constriction	Thoracic vertebral ganglia	Constriction and dilatation
Cerebral (p 339)	Dilatation	Superior cervical ganglion	Constriction
Abdominal and pelvic viscera (p 1093)	—	Prevertebral ganglia	Constriction
External genitalia (p 278)	Dilatation	Prevertebral ganglia	Constriction
<i>Eye</i>			
Iris (p 1174)	Constriction	Superior cervical ganglion	Dilatation
Ciliary muscle (p 1153)	Contraction	Superior cervical ganglion	Relaxation
Smooth muscle of orbit and upper lid (p 1181)	—	Superior cervical ganglion	Contraction
Nictitating membrane (cat and dog)	—	Superior cervical ganglion	Retraction
<i>Bronchi</i>	Constriction	Thoracic ganglia	Dilatation
<i>Glands</i>			
Sweat (p 734)	—	Vertebral ganglia	Secretion
Salivary (p 489)	Secretion	Superior cervical ganglia	Secretion
Gastric (p 505)	Secretion	Celiac ganglion	Inhibition?
Pancreas (p 531)			Secretion of mucus
Acini	Secretion	Celiac ganglion	—
Islets	Secretion	Celiac ganglion	—
Liver (p 535)	—	Celiac ganglion	Glycogenolysis
Adrenal (p 827) medulla	—	No postganglionic fibers	Secretion
<i>Smooth muscle</i>			
Of skin	—	Vertebral ganglia	Contraction
Of stomach wall (p 568)	Contraction or inhibition	Celiac ganglion	Contraction or inhibition
Of small intestine (p 583)	Increased tone and motility	Celiac and superior mesenteric ganglia	Inhibition
Of large intestine (p 584)	Increased tone and motility	Inferior mesenteric and hypogastric ganglia	Inhibition
Of bladder wall (p 483) (detrusor muscle)	Contraction	Inferior mesenteric and hypogastric ganglia	Inhibition
Of trigone and sphincter	Inhibition	Inferior mesenteric and hypogastric ganglia	Contraction
Of uterus, pregnant	nil	Inferior mesenteric and hypogastric ganglia	Contraction
Of uterus, non-pregnant	nil		Inhibition

* With certain exceptions, e.g., those supplying the sublingual and parotid glands and the sphincter pupillae, the postganglionic fibers of the parasympathetic arise from cells situated in, or in close proximity to, the innervated organ itself

an increase of circulating red cells does not occur during excitement or exercise (p 70), the usual reactions to cold (elevation of the hairs and vaso-

constriction) fail, and adrenaline is not liberated in an emergency. They are also less able to withstand oxygen lack or hemorrhage than are normal

animals. It is evident that the sympathectomized animal could not fend for itself, and, in the struggle for existence, would soon succumb to the hazards of the environment.

Parasympathetic effects, rather than being characterized by a diffuse outburst of activity, as may result from sympathetic stimulation, are more localized in character. It has also been suggested that they are concerned with conservative and restorative processes, and the sympathetic with processes involving the expenditure of energy. Inhibition of the heart, contraction of the pupil for the protection of the eye from intense light and the activities of the digestive tract, through which the energy stores of the body are replenished, are frequently given as examples of acts of conservation presided over by the parasympathetic. These apparent differences in the activities of the two divisions have led some (following Gaskell) to speak of the functions of the sympathetic and parasympathetic divisions as *catabolic* and *anabolic*, respectively. It is an interesting but perhaps a too speculative generalization.

A summary of the actions of the autonomic system upon various structures is given in table 98.

THE ACTIONS OF CERTAIN DRUGS UPON THE AUTONOMIC SYSTEM (see also p. 247). SYNAPTIC AND NEUROMUSCULAR TRANSMISSION, CHOLINESTERASE

Nicotine paralyzes ganglion cells or the junction (synapse) between the cell and the preganglionic fiber. The fibers themselves remain unaffected by the drug. An excitatory effect precedes the paralysis. The power of nicotine to paralyze ganglion cells renders it a most valuable means of locating the ganglion cells of autonomic nerves. If, for example, after a certain ganglion has been painted with the drug, stimulation of the fibers on the proximal side of the ganglion fails to cause the customary response, then it is concluded that the fibers in question have their cell stations in the treated ganglion. Nicotine also causes the contraction of striated bird and frog muscle and of denervated mammalian muscle. This excitatory effect is abolished by an excess of nicotine.

Curare. This drug blocks neuromuscular and synaptic transmission by antagonizing the action of acetylcholine. The researches of Cowan indicate that the antagonism is due to curare preventing the depolarization of the nerve or muscle by the choline ester (see below).

Adrenaline (ch. 59) exerts its action upon struc-

tures which are innervated by the sympathetic. The sweat glands of most animals and the adrenal medulla itself are exceptions. The responses which adrenaline evokes are similar to those caused by stimulation of the corresponding sympathetic nerves. Denervation of the effector organ increases its sensitivity to adrenaline. Burn and Bulbring have shown that adrenaline in low concentration has an augmenting effect upon the action of acetylcholine on junctional transmission (ganglionic and central synapses, and neuromuscular junctions), an effect contrary to that which it exhibits elsewhere, e.g., heart, intestine, etc. In high concentrations, on the other hand, adrenaline depresses the action of acetylcholine in these situations. Bulbring observed further that adrenaline is liberated from sympathetic ganglia during stimulation of the preganglionic fibers. This suggests that the enhancement of acetylcholine action by adrenaline is a physiological effect.

Ergotoxine and *ergotamine* paralyze motor and secretory fibers of the sympathetic but not the inhibitory fibers. When injected intravenously into the conscious animal (cat) ergotoxine produces a condition of "sham rage" (pp. 832 and 1027).

Pilocarpine causes parasympathetic effects, e.g., cardiac inhibition, contraction of the smooth muscle of the eye, bronchioles and alimentary tract, and secretion from the salivary, bronchial and gastric glands. It also stimulates the sweat glands which receive excitatory fibers from the sympathetic. The action of pilocarpine like that of muscarine or of acetylcholine is directly upon the effector cell.

Muscarine acts similarly to pilocarpine upon the heart and the smooth muscle of the alimentary tract, bladder and bronchioles. It constricts the pupil, and stimulates the salivary glands and sweat glands. It also causes vasodilatation.

Choline and its ester, *acetylcholine*, imitate closely the effects of parasympathetic stimulation, causing cardiac inhibition, excitation of the smooth muscle of the digestive tract and bladder wall, and the secretion of saliva and tears. It causes in addition dilatation of the arterioles, sweating and a fall in blood pressure. Acetylcholine is some 1000 times more powerful than choline itself. One part in many millions causes inhibition of the perfused frog's heart or contraction of an isolated intestinal segment. The ester is rapidly hydrolyzed in alkaline media into acetic acid and choline, and blood and other body fluids contain an enzyme—*cholinesterase*—which rapidly inactivates it. In ad-

dition to the parasympathetic effects of acetylcholine mentioned above, which have been termed its "muscarine" action, it has a stimulant action upon ganglion cells, upon the muscles of the body wall of the leech, upon voluntary frog muscle and upon denervated mammalian muscle. These effects of acetylcholine resemble those caused by nicotine and are referred to as its "nicotine" action.³ The latter, unlike the "muscarine" action, is not annulled by atropine but is abolished by a large dose of nicotine and by curare. Cholinesterase, specific for choline esters, is widely distributed throughout animal tissues, but is in greatest amounts in muscle, nerves and brain. In muscle, it is concentrated in the neighborhood of the motor end plates. In nervous tissue it is present in the nerve fiber, as well as in the cell body. It is present in the blood and, in man, is contained mainly in the red cells. The presence in serum of a *non-specific* cholinesterase was demonstrated by Mendel and Rudney. They named it *pseudocholinesterase*. It differs from *true* or *specific* cholinesterase in being capable of hydrolyzing non-choline esters, e.g., tributyrin, as well as acetylcholine. It is also relatively insensitive to the inhibitory action of eserine (see below), but is inactivated more readily by di-isopropylfluorophosphate than is true cholinesterase. A further dissimilarity between the two enzymes is with regard to the concentrations of acetylcholine at which they inactivate this substance. Pseudocholinesterase is most active when the acetylcholine concentration is high, whereas it has but slight hydrolytic action at the low (physiological) concentrations of acetylcholine at which true cholinesterase is most active. Pseudocholinesterase is found in pancreatic tissue as well as in serum, its function is unknown. Only true cholinesterase is present in nerve fibers and brain.

Anticholinesterases, *physostigmine* or *eserine*, *prosthigmine*, *neostigmine* and *di-isopropylfluorophosphate* (DFP). These drugs exert an action similar to that of acetylcholine because they destroy or inactivate cholinesterase. The actions of acetylcholine produced in the tissues (e.g., and muscle) is not effectively removed; its action is thus prolonged and intensified. Di-isopropylfluorophosphate (DFP) is an especially powerful anticholinesterase. Its administration is followed by the effects of severe acetylcholine poisoning. Later,

³ The muscarine action of acetylcholine is exerted on the heart muscle, secreting glands and smooth muscle, whereas its effect upon the cells of the adrenal medulla, muscle end plates, and the cells of sympathetic and parasympathetic ganglia is a nicotine action.

when the cholinesterase falls to a low level, nerve conduction fails, owing, apparently, to failure of the nerve to become repolarized (p. 926). The inactivation of cholinesterase by DFP is irreversible, whereas that caused by the alkaloids, eserine, prostigmine or neostigmine, is reversible. Prostigmine also differs in its action from both eserine and DFP, for the latter two drugs abolish nerve conduction, prostigmine does not. It has been suggested that this difference is due to the prostigmine being unable to penetrate the nerve membrane and thus to come into contact with the enzyme. Tetraethylpyrophosphate (TEPP) has an action similar to that of DFP.

Atropine annuls parasympathetic effects, and antagonizes the actions of pilocarpine, of muscarine and of the "muscarine" action of acetylcholine. It is believed to act directly upon the effector cell, preventing the action but not the liberation of acetylcholine. It therefore quickens the heart, and paralyzes the sphincter pupillae and ciliary muscles and thus causes pupillary dilatation and failure of accommodation. It suppresses the secretion of saliva, and of the nasal, bronchial and gastric glands. It is inhibitory to the bronchial muscles and lowers the tone of the intestinal musculature. *It also suppresses the secretion of the sweat glands* which are innervated by the sympathetic.

THE CHEMICAL TRANSMISSION OF NERVOUS EFFECTS TO AUTONOMIC EFFECTORS AND SKELETAL MUSCLE

The liberation of chemical substances (humors) by nerves

The work of a number of investigators, starting with that of Loewi, has fully established the fact that autonomic effects are associated with the liberation of chemical substances at the nerve terminals. In the case of parasympathetic fibers (preganglionic and postganglionic) the chemical is acetylcholine. This substance is also produced by preganglionic sympathetic nerves. Adrenaline or noradrenaline liberation (ch. 59) occurs with the arrival of the impulse at postganglionic sympathetic terminals.

Since anatomical terms fail to express these physiological conceptions, Dale suggested that the fibers which liberate acetylcholine be called *cholinergic*, and the term *adrenergic* be used for those which liberate an adrenaline-like substance. Thus (a) parasympathetic postganglionic fibers, (b)

sympathetic postganglionic fibers supplying the sweat glands and uterus, and (c) sympathetic preganglionic fibers, are cholinergic—the “muscarine” action being prominent in the first two instances, the “nicotine” action in the last (d) The parasympathetic preganglionic fibers and the postganglionic sympathetic fibers causing vasodilatation (see p 1101) are also cholinergic. Postganglionic sympathetic fibers, e.g., to the heart, intestine, etc., as well as those causing vasoconstriction, are adrenergic (fig 724)

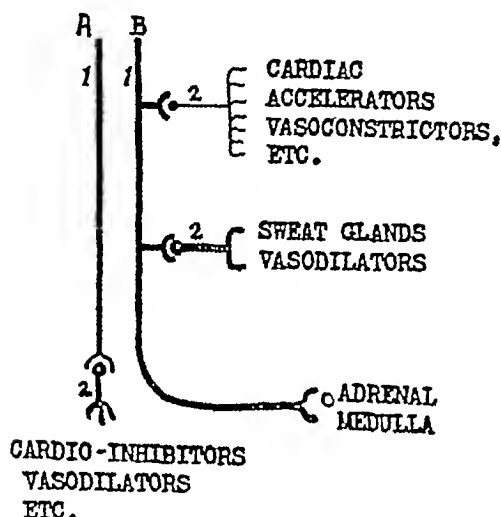


FIG 724 Diagram showing the distribution of cholinergic and adrenergic fibers A, parasympathetic, B sympathetic 1 preganglionic, 2, postganglionic, heavy lines, cholinergic, light lines, adrenergic.

Acetylcholine is present in the tissues in an inactive bound form (probably in loose combination with protein or lipoprotein), and is active only after it has been freed also in this form alone is it susceptible to the action of cholinesterase. While bound it is protected from the action of the enzyme.

Space does not permit an extended account of the experimental work upon which present day knowledge of the rôle played by these substances is based, but a summary of the more important facts will be attempted. The work of Loewi upon the transmission of cardiac effects (vagal and accelerator) and Cannon's discovery of “sympathin” have already been considered in ch. 59⁴

The experiments to be described involve either (a) a comparison of the effects of autonomic and

⁴ The suggestion that a chemical substance might be liberated from nerve endings was made by Elliott (1904-5), and before him by Du Bois Reymond, though no direct experimental evidence was offered

motor nerve stimulation with the pharmacological actions of acetylcholine or, (b) the stimulation of the autonomic nerves to an organ, e.g., the salivary gland, and testing the venous blood issuing from it, or the fluid with which it has been perfused, for an acetylcholine-like action. In order to prevent the hydrolysis by cholinesterase of acetylcholine after its liberation from the nerve terminals, eserine in a dilution of 1 in 1 million or so is added to the perfusion fluid. The following is a list of the tests employed and the pharmacological effects which indicate the presence of acetylcholine.

(1) *Blood pressure of the cat* Dilatation of the arterioles and a fall in blood pressure (“muscarine” action). The effect should be annulled by atropine. (2) *Inhibition of the perfused frog's heart or rabbit's auricle* (“muscarine” action annulled by atropine). (3) *Contraction of the voluntary muscle (rectus abdominis) of the frog* (“nicotine” action). (4) *Contraction of the muscle in body wall of the eschmized leech* (“nicotine” action). This test is sensitive to 2 gamma of acetylcholine per liter. (5) The active substance should be inactivated by alkali or by uneserized blood (which contains cholinesterase).

During nerve stimulation the acetylcholine-like substance is liberated and enters the blood or perfusion fluid in such minute amounts that its chemical identification is out of the question. Had acetylcholine not been demonstrated by chemical analysis to be a normal constituent of the body, such purely pharmacological tests for its appearance during nerve stimulation would still leave some doubt that it was actually the substance concerned. This doubt was removed by the work of Dale and Dudley, who obtained chemically recognizable amounts of this choline ester from the spleen of the horse. It is also present in human placenta and in nervous tissues, only minute amounts are present in blood. That the chemical mediator is actually acetylcholine is now generally accepted. With one exception (namely pyruvylcholine) no choline ester other than acetylcholine has an action quantitatively comparable with that exhibited by the chemical transmitter. Moreover, acetylcholine is the only ester of choline found in animal tissues.

Summary of experimental work relating to the rôle played by acetylcholine in the transmission of nervous effects

(1) *Stimulation of the chorda tympani to salivary glands* Babin and his associates found that when the chorda of one side was stimulated a substance entered

the blood which caused a fall in blood pressure and secretion from the denervated salivary gland of the opposite side. These effects were abolished by atropine. Similar results have been obtained by others. During stimulation of the chorda of a perfused salivary gland, for example, a substance identical in action with acetylcholine was found by Henderson and Roepke in the perfusion fluid.

(2) *The liberation of acetylcholine from parasympathetic endings in the iris.* When a strong light was thrown into one eye, the other eye being shaded, an acetylcholine-like substance was obtained from the aqueous humor of the illuminated eye but not from the darkened eye (Engelhart).

(3) *Acetylcholine and vasodilator nerves. The hypersensitivity of denervated structures to chemical stimulation.* Many years ago Philipeaux and Vulpian (1863) observed that after section and degeneration of the hypoglossal nerve, stimulation of the chorda tympani going to the tongue caused a peculiar slow and prolonged contraction of the lingual muscles (Vulpian effect) and vasodilatation. Stimulation of the chorda causes no effect upon the normal tongue other than vasodilatation. Acetylcholine also causes this peculiar contraction of the denervated tongue muscles ("nicotine" action) together with vasodilatation ("muscarine" action). It has therefore been suggested that the vasodilatation following chorda stimulation is due to the liberation of acetylcholine from the nerve terminals, and that the Vulpian effect results from the diffusion of the active substance to the muscle fibers rendered sensitive by denervation. A reaction analogous to the Vulpian effect was described by Sherrington, who found that after the muscles of the leg were deprived of their motor innervation by sectioning the ventral roots containing fibers for the sciatic nerve, and allowing time for the degeneration of the fibers to occur, stimulation of the sciatic caused the characteristic slow contraction of the muscles. The effect can be duplicated by acetylcholine. Evidence was subsequently obtained which indicated that the nerve fibers concerned ended upon the blood vessels. It was therefore concluded that the effect was due to the liberation of acetylcholine from the sensory fibers which normally carried antidromic vasodilator impulses (p. 276). It now appears, however, from the work of Hinsey and Cutting, that sympathetic postganglionic fibers arising from the lumbar and sacral ganglia and reaching the sciatic nerve via the gray rami are responsible. This leads to the suggestion that acetylcholine liberation from sympathetic terminals upon the blood vessels causes vasodilatation and the Sherrington phenomenon as a secondary effect due to diffusion of the ester from its site of production to the muscle fibers. The latter as well as the Vulpian effect is intensified by eserine. An effect of the same nature as the Sherrington phenomenon is obtained upon stimulation of the cervical sympathetic after section and degeneration of the facial

nerve, viz., dilatation of the vessels of the gums and lips together with contraction of the muscles of the upper lip (Rogowitz).

Another example of increased sensitivity of denervated muscle to acetylcholine is the interesting and suggestive observation made more recently by Bender. After section and degeneration of the 7th nerve in monkeys the denervated facial muscles were found to contract involuntarily when the animal became angry or frightened. This "fright reaction" can be duplicated by the injection of acetylcholine and is accentuated by eserine. It is attributed to the release of acetylcholine into the general circulation from some unknown source. The phenomenon recalls the liberation of adrenaline in emotional states and the hypersensitivity to adrenaline or sympathin of the denervated iris. Further examples of the hypersensitivity of denervated structures to acetylcholine, adrenaline and other chemical agents can be cited, e.g., the greater secretory response of the denervated salivary gland to pilocarpine, the hypersensitivity of the pupil to acetylcholine after section of the 3rd nerve and of the nictitating membrane to adrenaline following excision of the superior cervical ganglion. The inhibitory effects of adrenaline (e.g., on the bowel and non-pregnant uterus) are also enhanced by denervation. These and other similar observations have led to the formulation by Cannon of a *law of denervation* which he states in the following terms: "When in a series of efferent neurons a unit is destroyed, increased irritability to chemical agents develops in the isolated structure or structures, the effect being maximal in the part directly denervated."

If the effects described in the first paragraph of this subsection are actually due to acetylcholine one would expect the vasodilatation ("muscarine" action) to be prevented by atropine. But this is not the case. Though atropine annuls the vasodilator action of acetylcholine it does not prevent the effects of stimulating vasodilator nerves. It has been suggested (Dale), in an effort to explain this discordant fact, that acetylcholine is liberated in such intimate relation to the smooth muscle of the vessel that atropine fails to reach it.

(4) *Acetylcholine as an intermediary of parasympathetic effects to the alimentary tract and bladder.* It has been shown that Ringer's solution in which a beating loop of intestine is immersed is capable of augmenting the activity of another similar loop (Weiland). It has also been found that if the vagus nerve to the intestine is stimulated and a loop then removed and suspended in Ringer's solution, the contractions of this loop are greater than those of a similar one which had not been previously excited in this manner. These experiments suggest that during intestinal activity a substance is liberated by the vagal endings which has an augmenting effect upon the contractions. Evidence for the liberation of acetylcholine by the gastric vagus has recently been obtained by Dale and Feldberg. A substance identical in action with acetylcholine was detected in the venous

blood leaving the resting stomach or in the esophageal fund perfusing its wall. During vagal stimulation the quantity of the secreted material was increased fourfold.

An acetylcholine-like substance has also been identified by Henderson and Rooke in the fund perfusing the duodenum during stimulation of its parasympathetic nerves. In the case of the intestine and duodenum as in the case of vasodilator nerves the same discrepancy exists between the action of atropine upon the effects of nerve stimulation and the action of the drug upon the effects of acetylcholine administration. Atropine abolishes the action of acetylcholine when applied artificially to these organs but does not abolish the contractions set up by parasympathetic stimulation. Henderson and Rooke conclude from the results of their experiments that whereas the tone of the intestine and of the bladder is dependent upon the liberation of acetylcholine, another mechanism is responsible for the passive contractions of intestine as is well known, decreases the tone of the intestinal and vesical musculature but exerts no direct effect upon the contractile mechanism.

5. *The liberation of acetylcholine from sympathetic ganglia.*—Dale, Feldberg and Geerup perfused the superior cervical ganglion with esophageal fund. The fundus muscle was inserted into the common carotid artery all branches of which had been cut except the one to the ganglion. The fluid was collected from the internal jugular vein, all its tributaries except that from the ganglion having been occluded. During stimulation of the cervical sympathetic trunk below the ganglion the fluid arising from the vein was found to possess an action identical with that of acetylcholine. When the collected fluid was passed through the ganglion of the opposite side stimulation ("transmission") action up to a certain structure was evidenced by a contraction of the contracting structure. Fluid collected before or after the period of stimulation showed no such action.

6. *Source of liberation during the discharge of adrenaline.*—Stimulation of the sympathetic fibers supplying the adrenal medulla causes an acetylcholine-like substance to appear in the venous blood of the adrenal vein. It has therefore been concluded that acetylcholine acts as a chemical transmitter from the nerve terminals to the medullary cells. It will be recalled that the sympathetic fibers ending in the adrenal are preganglionic, the adrenal cell itself taking the place of the ganglion cell (p. 828). It is mainly the "muscarinic" action of acetylcholine like, abolished by a large dose of muscarine but not by atropine, which is exerted upon the adrenal, but there is also a slight "transmission" action.

7. *Effect of the action of transmitter of fluids to the rest of the body.*—The organs of the human sweat glands are innervated by the sympathetic fibers. The action of drugs is similar to that of structures supplied by the parasympathetic, they are unaffected by adrenaline, stimulated

by pilocarpine, and paralyzed by atropine. An experiment by Dale and Feldberg gives an explanation of these discrepancies, or at any rate brings the mechanism of sweat secretion into the general scheme. Excitation of the sympathetic fibers to the foot pads of the cat was followed by sweating and the appearance of acetylcholine in the esophageal fund perfusing the paw.

(8) *Control transmission of motor nerve impulses to voluntary muscle.*—Evidence that acetylcholine is liberated from the terminals of motor nerves and serves as a transmitter of impulses to the muscle fibers has been obtained by Dale and his associates.

(a) Upon rhythmic stimulation of the hypoglossal nerve of the perfused tongue of the cat, acetylcholine appeared in the venous fund. Similar results were secured with the perfused leg muscles of the dog during stimulation of the ventral spinal roots after excision of the lumbar sympathetic chain.

(b) The sudden injection of a small dose (2 to 10 γ) of acetylcholine into the artery supplying the gastrocnemius of the cat during circulatory arrest caused a sharp contraction of the muscle. This is a "muscarinic action" of acetylcholine, it is abolished by curarine but not by atropine. The direct application of a minute amount ($5 \times 10^{-4} \gamma$) of acetylcholine to the motor end-plate of the muscle fiber causes a short sharp tetanic contraction. Ten times this quantity applied elsewhere to the fiber is ineffective (Büchthal and Lindhard).

(c) The muscular administration of eserine (0.2 to 0.3 mg. per kg.) on a small cat caused an increase of 150 per cent in tension of a gastrocnemius twitches provoked by stimulation of the motor nerve. Eserine had no such effect upon the response of denervated muscle to direct stimulation.

(9) There is evidence for the cholinergic nature of the secretory nerves of the neural lobe of the hypophysis.

From the results of the investigations cited in the foregoing paragraphs, it can be stated that acetylcholine production and liberation accompany the excitation of parasympathetic nerves, such as the chorda tympani going to the salivary glands, and those to the sweat glands, adrenal medulla, neural lobe of the pituitary, sphincter pupillae, heart, stomach, bronchioles, erectile tissue, voluntary muscle, etc. It is also probably liberated by both sympathetic and parasympathetic vasodilator nerves and by both types of nerves supplying the intestine and bladder. All preganglionic fibers are cholinergic.

The role played by potassium ions in synaptic transmission.

The injection of small amounts of potassium chloride into the fund perfusing the superior cervical ganglion stimulates the nerve cells, and the injection of a sub-

threshold dose augments the effect of a series of sub-maximal stimuli applied to the preganglionic fibers (Feldberg and Vartiainen). When the concentration of potassium chloride is raised to four times the normal, acetylcholine appears in the venous outflow from the ganglion (Feldberg and Brown). It had been reported previously by Beznak that KCl caused the liberation of acetylcholine in the frog's heart. After section of the cervical sympathetic and degeneration of the preganglionic fibers the addition of KCl to the fluid perfusing the superior cervical ganglion, though still causing excitation of the cells, results in the liberation of an insignificant amount of acetylcholine. This fact indicates that the K ion itself (i.e., quite independent of acetylcholine) has a stimulant action upon the ganglion cells. There has been some discussion of the interpretation of these findings in respect to the rôle played by the K ion in the normal conduction in sympathetic ganglia. Brown and Feldberg found, for example, that curarine, which blocks transmission at synaptic junctions, does so by preventing acetylcholine from acting on the ganglion cell. The liberation of acetylcholine at the preganglionic ending is not affected by curarine. These observations suggest that potassium is not directly concerned in the transmission of the effect across the synapse. Cowan's observations suggest that the effect of curare upon the myoneural junction is due to its preventing the depolarization of the end-plate caused by acetylcholine.

A SUMMARY OF THE THEORIES OF SYNAPTIC AND NEUROMUSCULAR TRANSMISSION

According to the original views of Dale and his associates, acetylcholine was supposed to act as a direct stimulant to the ganglion cell or the muscle end-plate (see fig. 72.5). That is, the nerve impulse upon arriving at the preganglionic terminal or nerve-ending in the muscle end-plate, liberated acetylcholine which acted as a chemical transmitter of the nervous effect to the postjunctional structure, a fresh impulse being thus set up.

Since it is known that each impulse in the preganglionic fiber causes only one impulse to be discharged from the ganglion cell, it follows that the acetylcholine liberated must be immediately destroyed by the action of cholinesterase and the nerve fiber restored to its original state. That is, the removal of acetylcholine is necessary for the recovery of the nerve. Thus, in brief, is the *chemical theory* of junctional transmission.

As a result of much research on this subject in recent years the *chemical theory* in the simple form outlined above has given place to other views. One of the first objections advanced (Eccles) to the chemical theory as applied to rapid types of nervous action, namely, the

transmission of excitation at a synapse or myoneural junction ("nicotine" action), was the great rapidity with which acetylcholine must be removed after it had exerted its effect. It must disappear within the refractory period of nerve—a matter of milliseconds at most. This objection need not apply to the slower types of nervous activity, such as the inhibitory effect of the vagus on the heart ("muscarine" action). It was also doubted that sufficient quantities of acetylcholine for excitation could be produced with such lightning-like rapidity. A still more serious objection was the fact that only relatively small amounts of acetylcholine could be recovered from fluid perfusing a ganglion during excitation of the preganglionic fiber. Even though eserine is employed to prevent the destruction of acetylcholine, only about 1/40,000 of the quantity required to stimulate the ganglion cell can be obtained, and only 1/100,000 of the quantity necessary to stimu-

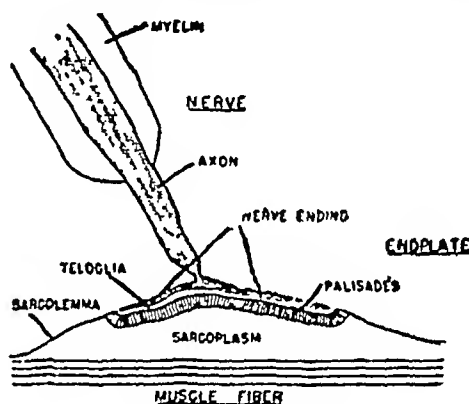


FIG. 72.5 Diagram of the region of the muscle end-plate (from Acheson after Couteaux)

late muscle. The experiments of Marnay and Nachmansohn have answered some of these objections. They have furnished evidence of a high concentration of cholinesterase at the motor end-plates. According to their estimation, the concentration of enzyme at these sites is many thousand times greater than in the rest of the muscle.

The rate of production of the enzyme at the muscle end-plates is the basis for the estimation of the formation of its substrate, acetylcholine. In other words, the rate at which acetylcholine is hydrolysed is taken as a measure of its production. In this way it has been estimated that about 2×10^{-6} micrograms of acetylcholine can be split at a single nerve ending of the frog's sartorius muscle within the time of the refractory period. This amount which corresponds to the production of 8×10^9 molecules of acetylcholine, is considered to be more than sufficient to stimulate. It has also been found that in a sympathetic ganglion from 3 to 6×10^{12} molecules of the ester can be hydrolyzed within a millisecond.

The discovery of an enzyme in extracts of brain and nerve which synthesizes acetylcholine at a high rate

blood leaving the resting stomach or in the eserinizd fluid perfusing its wall. During vagal stimulation the quantity of the active material was increased four-fold.

An acetylcholine like substance has also been identified by Henderson and Roepke in the fluid perfusing the bladder during stimulation of its parasympathetic nerves. In the case of the intestine and bladder as in the case of vasodilator nerves, the same discrepancy exists between the action of atropine upon the effects of nerve stimulation and the action of the drug upon the effects of acetylcholine administration. Atropine abolishes the action of acetylcholine when applied artificially to these organs but does not depress the contractions set up by parasympathetic stimulation. Henderson and Roepke conclude from the results of their experiments that, whereas the tone of the intestine and of the bladder is dependent upon the liberation of a choline ester, another mechanism is responsible for the phasic contractions. Atropine, as is well known, depresses the tone of the intestinal and vesical musculatures but exerts no direct effect upon the contractile mechanism.

(5) *The liberation of acetylcholine from sympathetic preganglionic fibers.* Feldberg and Gaddum perfused the superior cervical ganglion with eserinizd fluid. The inflow cannula was inserted into the common carotid artery, all branches of which had been tied except the one to the ganglion. The fluid was collected from the internal jugular vein, all its tributaries except that from the ganglion having been occluded. During stimulation of the cervical sympathetic trunk below the ganglion the fluid issuing from the vein was found to possess an action identical with that of acetylcholine. When the collected fluid was passed through the ganglion of the opposite side its stimulant action ("nicotine" action) upon this structure was evidenced by a contraction of the mictitating membrane. Fluid collected before or after the period of stimulation showed no such activity.

(6) *Acetylcholine liberation during the discharge of adrenaline.* Stimulation of the sympathetic fibers supplying the adrenal medulla causes an acetylcholine like substance to appear in the blood of the adrenal vein. It has therefore been concluded that acetylcholine acts as a chemical transmitter from the nerve terminals to the medullary cells. It will be recalled that the sympathetic fibers ending in the adrenal are preganglionic, the adrenal cell itself taking the place of the ganglion cell (p. 828). It is mainly the "nicotine" action of acetylcholine (i.e., abolished by a large dose of nicotine but not by atropine) which is exerted upon the adrenal, but there is also a slight "muscarine" action.

(7) *Acetylcholine as the transmitter of effects to the sweat glands.* Though the human sweat glands are innervated by the sympathetic their behavior to drugs is similar to that of structures supplied by the parasympathetic, they are unaffected by adrenaline, stimulated

by pilocarpine, and paralyzed by atropine. An experiment by Dale and Feldberg gives an explanation of these discrepancies, or at any rate brings the mechanism of sweat secretion into the general scheme. Excitation of the sympathetic fibers to the foot pads of the cat was followed by sweating and the appearance of acetylcholine in the eserinizd fluid perfusing the paw.

(8) *Chemical transmission of motor nerve impulses to voluntary muscle.* Evidence that acetylcholine is liberated from the terminals of motor nerves and serves as a transmitter of impulses to the muscle fibers has been obtained by Dale and his associates.

(a) Upon rhythmic stimulation of the hypoglossal nerve of the perfused tongue of the cat, acetylcholine appeared consistently in the venous fluid. Similar results were secured with the perfused leg muscles of the dog during stimulation of the ventral spinal roots after excision of the lumbar sympathetic chain.

(b) The sudden injection of a small dose (2 to 10 γ) of acetylcholine into the artery supplying the gastrocnemius of the cat during circulatory arrest caused a sharp contraction of the muscle. This is a "nicotine action" of acetylcholine, it is annulled by curarine but not by atropine. The direct application of a minute amount ($5 \times 10^{-8} \gamma$) of acetylcholine to the motor end plate of the muscle fiber causes a short sharp tetanic contraction. Ten times this quantity applied elsewhere to the fiber is ineffective (Buchthal and Lindhard).

(c) The intravenous administration of eserine (0.2 to 0.3 mg per kg) to a spinal cat caused an increase of 130 per cent in tension of a gastrocnemius twitch provoked by stimulation of the motor nerve. Eserine had no such effect upon the response of denervated muscle to direct stimulation.

(9) There is evidence for the cholinergic nature of the secretory nerves of the neural lobe of the hypophysis.

From the results of the investigations cited in the foregoing paragraphs, it can be stated that acetylcholine production and liberation accompany the excitation of parasympathetic nerves, such as the chorda tympani going to the salivary glands, and those to the sweat glands, adrenal medulla, neural lobe of the pituitary, sphincter pupillae, heart, stomach, bronchioles, erectile tissue, voluntary muscle, etc. It is also probably liberated by both sympathetic and parasympathetic vasodilator nerves, and by both types of nerves supplying the intestine and bladder. All preganglionic fibers are cholinergic.

The rôle played by potassium ions in synaptic transmission

The injection of small amounts of potassium chloride into the fluid perfusing the superior cervical ganglion stimulates the nerve cells, and the injection of a sub-

threshold dose augments the effect of a series of submaximal stimuli applied to the preganglionic fibers (Feldberg and Vartiainen). When the concentration of potassium chloride is raised to four times the normal, acetylcholine appears in the venous outflow from the ganglion (Feldberg and Brown). It had been reported previously by Beznak that KCl caused the liberation of acetylcholine in the frog's heart. After section of the cervical sympathetic and degeneration of the preganglionic fibers the addition of KCl to the fluid perfusing the superior cervical ganglion, though still causing excitation of the cells, results in the liberation of an insignificant amount of acetylcholine. This fact indicates that the K ion itself (i.e., quite independent of acetylcholine) has a stimulant action upon the ganglion cells. There has been some discussion of the interpretation of these findings in respect to the rôle played by the K ion in the normal conduction in sympathetic ganglia. Brown and Feldberg found, for example, that curarine, which blocks transmission at synaptic junctions, does so by preventing acetylcholine from acting on the ganglion cell. The liberation of acetylcholine at the preganglionic ending is not affected by curarine. These observations suggest that potassium is not directly concerned in the transmission of the effect across the synapse. Cowan's observations suggest that the effect of curare upon the myoneural junction is due to its preventing the depolarization of the end-plate caused by acetylcholine.

A SUMMARY OF THE THEORIES OF SYNAPTIC AND NEUROMUSCULAR TRANSMISSION

According to the original views of Dale and his associates, acetylcholine was supposed to act as a direct stimulant to the ganglion cell or the muscle end-plate (see fig. 72.5). That is, the nerve impulse upon arriving at the preganglionic terminal or nerve-ending in the muscle end-plate, liberated acetylcholine which acted as a chemical transmitter of the nervous effect to the postjunctional structure, a fresh impulse being thus set up.

Since it is known that each impulse in the preganglionic fiber causes only one impulse to be discharged from the ganglion cell, it follows that the acetylcholine liberated must be immediately destroyed by the action of cholinesterase and the nerve fiber restored to its original state. That is, the removal of acetylcholine is necessary for the recovery of the nerve. This, in brief, is the *chemical theory* of junctional transmission.

As a result of much research on this subject in recent years the chemical theory in the simple form outlined above has given place to other views. One of the first objections advanced (Eccles) to the chemical theory as applied to rapid types of nervous action, namely, the

transmission of excitation at a synapse or myoneural junction ("nicotine" action), was the great rapidity with which acetylcholine must be removed after it had exerted its effect. It must disappear within the refractory period of nerve—a matter of milliseconds at most. This objection need not apply to the slower types of nervous activity, such as the inhibitory effect of the vagus on the heart ("muscarine" action). It was also doubted that sufficient quantities of acetylcholine for excitation could be produced with such lightning-like rapidity. A still more serious objection was the fact that only relatively small amounts of acetylcholine could be recovered from fluid perfusing a ganglion during excitation of the preganglionic fiber. Even though eserine is employed to prevent the destruction of acetylcholine, only about 1/40,000 of the quantity required to stimulate the ganglion cell can be obtained, and only 1/100,000 of the quantity necessary to stimu-

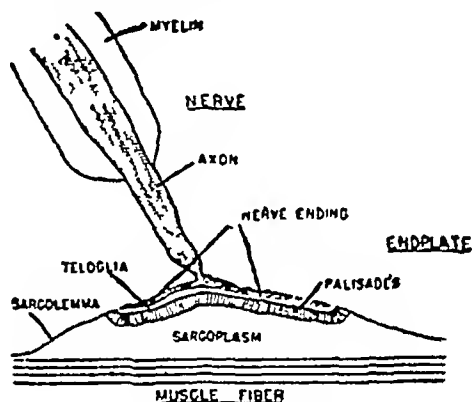


FIG. 72.5 Diagram of the region of the muscle end-plate (from Acheson after Couteaux)

late muscle. The experiments of Marnay and Nachmansohn have answered some of these objections. They have furnished evidence of a high concentration of cholinesterase at the motor end-plates. According to their estimation, the concentration of enzyme at these sites is many thousand times greater than in the rest of the muscle.

The rate of production of the enzyme at the muscle end-plates is the basis for the estimation of the formation of its substrate, acetylcholine. In other words, the rate at which acetylcholine is hydrolysed is taken as a measure of its production. In this way it has been estimated that about 2×10^{-6} micrograms of acetylcholine can be split at a single nerve ending of the frog's sartorius muscle within the time of the refractory period. This amount which corresponds to the production of 8×10^9 molecules of acetylcholine, is considered to be more than sufficient to stimulate. It has also been found that in a sympathetic ganglion from 3 to 6×10^{12} molecules of the ester can be hydrolyzed within a millisecond.

The discovery of an enzyme in extracts of brain and nerve which synthesizes acetylcholine at a high rate

under anaerobic conditions indicates that a mechanism exists for the rapid production of the choline ester. This enzyme is called *choline acetylase*.⁵ Nachmansohn and his associates suggest that the energy for the resynthesis of acetylcholine is derived from the break down of adenosinetriphosphate.

Those who supported the classical theory of neuromuscular and synaptic transmission, namely, that the action currents set up in the motor nerve or preganglionic fiber excite the muscle or the ganglion cells, point out that the interval between the arrival of the nerve impulse at the motor end plate and the beginning of the propagated disturbance in the muscle (end plate delay), is of about the same duration as the "spike" potential. Erlanger has shown also that the nerve impulse can excite the section of nerve beyond a blocked region of from 1 to 2 millimeters in length. He argued, therefore, that it was unreasonable to maintain that discontinuity at a synapse (or presumably at the myoneural junction) would prevent transmission of the impulse. Such facts advanced by the proponents of the electrical theory could not be disregarded, yet, it was difficult to believe that the liberation of acetylcholine by nerve impulses, a fact proved beyond dispute, and the relatively high concentration of cholinesterase at the precise site where it could play an essential rôle in the chemical mechanism, and the widespread distribution in nervous tissue of a specific enzyme for the synthesis of acetylcholine, are merely incidental and without any physiological significance.

Within more recent years attempts have been made, especially by Nachmansohn and his associates, to harmonize the chemical and electrical conceptions. Acetylcholine production by nerve and the electrical changes are considered by Nachmansohn as two inseparable parts of one process. The elaboration of the chemical is viewed as an essential accompaniment of the conduction of the impulse along the nerve fiber. It is liberated, not only at the nerve terminals but throughout the length of the axon. Its production is thought to be a primary event in excitation, preceding the electrical response.

The theory of Nachmansohn is based upon the classical membrane theory of excitation and conduction in nerve (p. 926), and the correlation between electrical and chemical phenomena (drug actions, cholinesterase concentrations, etc.).⁶ The theory, briefly stated, is

⁵ Choline acetylase is very widely distributed in the nervous tissues, being found in brain, ganglia and peripheral nerves, it is also present in the placenta and in electric organs, but is absent from liver and kidney. For optimal activity this enzyme requires Mg, K and Ca ions. The concentration of choline acetylase in nerve fibers declines during their degeneration, but some still remains even after the nerve has ceased to conduct.

⁶ The close relationship between acetylcholine metabolism and the E.M.F. of the action potential has been shown clearly by Nachmansohn and his associates in a study of the electric organs of such fish as the electric eel and torpedo. These organs can discharge an electric shock which in some of the larger species

as follows. The nerve fiber is surrounded by a semi-permeable membrane, polarized during the resting state, with potassium ions concentrated upon the inside. An increase in permeability of the membrane causes potassium ions to leak out, depolarization and reduced resistance of the nerve fiber result. The depolarized region is electrically negative to other parts of the nerve, a current is thus set up which is associated with the propagation of a wave of excitation along the nerve fiber (p. 926). The rôle of increasing membrane permeability and depolarization, or of lowering membrane resistance, is assigned to acetylcholine. The propagated electrical disturbance, which is accompanied by the release of acetylcholine, upon arriving at the synapse or myoneural junction, crosses the "gap" (see fig. 72.5) and excites the ganglion cell or the muscle fiber. Thus, the function of acetylcholine in transmission of the impulse is the same in the nerve fiber as across a junction. Having brought about depolarization, acetylcholine must be rapidly removed in order that repolarization shall occur and the nerve be ready to transmit another impulse. This "cocking of the trigger", as it were, is brought about by cholinesterase. A new supply of acetylcholine (probably by the esterification of choline present in the nerve sheath) must be produced. This is the function of choline acetylase. The energy for resynthesis is furnished by the phosphate rich bonds of adenosinetriphosphate.

There are several observations which have been raised against this theory, one of which is the very small amount of acetylcholine in sensory nerves. If acetylcholine were essential for conduction of the nerve impulse, one would expect to find comparable amounts of these substances in both types of nerve.

It has been even more difficult to obtain decisive evidence with respect to the rôle played by acetylcholine in the central nervous system, but whatever its rôle here may be, it is probably the same as that in ganglionic synapses. That its function is an important one is evidenced by several facts, but whether it is involved directly in synaptic transmission, is, as in the case of ganglionic synapses, a highly controversial question.

The following observations can be cited in favor of

amounts to from 400 to 800 volts. The electric organ is composed of plate like structures arranged in series or columns which are believed to have evolved from the end plates of skeletal muscle. They resemble in their arrangement a voltaic pile. One side only of each plate is innervated, and this side becomes negative during activity. The other side is positive. High concentrations of cholinesterase are produced in these organs. In some instances several kilograms are formed in an hour, or an amount three or more times the weight of the organ itself. A close correlation has been shown by Nachmansohn and his colleagues to exist in the electric organ of *Electrophorus electricus*, between the number of plates per cm., the E.M.F. per cm. and the concentration of cholinesterase. There remains little doubt that the production of acetylcholine is closely associated with the activation and electrical discharge of these organs.

the chemical theory of synaptic transmission in the central nervous system

(1) Activity in the central nervous system is continuous, and the continuous or "spontaneous" liberation of acetylcholine has been demonstrated repeatedly, cholinesterase and choline acetylase are present in brain in considerable amounts

(2) Eserine applied to various parts of the central nervous system has a stimulating action. The following responses have been obtained, (a) movements of the opposite limbs after the application of the enzyme to the motor area of the cortex, (b) limb movements after application to anterior lobe of the cerebellum, (c) movements of the tongue and of the muscles of deglutition when painted on the hypoglossal nucleus (see Miller), (d) liberation of the hormone of the posterior lobe of the pituitary after its application to the supraoptic nucleus

(3) Bulbring and Burn perfused the lower half of the spinal cords of dogs with an eseriniz solution, this was followed by activity of a reflex nature which persisted for several minutes, stimulation of the central end of the sciatic nerve during this period caused an increase in the amount of acetylcholine in the effluent fluid

Myasthenia gravis This condition, as its name implies, is a condition of profound weakness of the muscles, those of the eyes, face and throat being, as a rule, involved first. Collections of lymphocytes (lymphorrhages) and degenerative changes occur in the muscles. The disease is accompanied by a high degree of creatinuria and a reduction in the excretion of creatinine. The thymus is frequently enlarged, and myasthenia gravis is a frequent accompaniment of thymic tumors. Administered creatine is practically all excreted as such. Death results from involvement of the respiratory muscles. Myasthenia gravis is associated sometimes with thyrotoxicosis in which enlargement of the thymus is also commonly seen

In 1934 Walker described a dramatic improvement in a case of myasthenia gravis following the administration of eserine, later prostigmine was found to confer even greater benefit. The effect, unfortunately, is transient, lasting for no longer than 3 or 4 hours. Other anticholinesterases, e.g., neostigmine and di-isopropyl-fluorophosphate, have since been employed, the latter agent, though it has a more lasting effect is otherwise inferior to prostigmine or neostigmine. The first drug to be used with benefit was ephedrine, introduced by Edgeworth in 1930, who was herself a victim of the disease

The muscular weakness in myasthenia gravis is not due to any disease of the central nervous system, nerve trunks, or muscles but, as shown by Harvey and his

associates, to some defect at the myoneural junction. With regard to the nature of the junctional disorder, the action of cholinesterases suggests three possibilities, namely (a) deficiency of acetylcholine production, (b) the presence of excessive amounts of cholinesterase, or (c) the production of a substance with a curare-like action. The investigations of Wilson and Stoner point to the last mentioned as the most likely factor. They found no increase in the cholinesterase content of the blood in subjects of the disease, nor any evidence of a failure in acetylcholine synthesis, but the serum of patients suffering from the disease, but not under treatment with prostigmine, caused neuro-muscular block in the nerve-muscle preparation of the frog. Serum from patients receiving prostigmine treatment exerted no curare-like action. With the idea that the primary fault originates in the thymus, thymectomy has been resorted to in some instances, with apparently beneficial results

AFFECTIONS OF AUTONOMIC NERVES

HORNER'S SYNDROME is the name given to the group of effects resulting from section or paralysis of the cervical sympathetic. They are (a) *Ptoxis* (drooping of eyelid) and *enophthalmos* (recession of the eyeball) due to paralysis of the smooth muscle of the upper lid and orbit respectively, reduction in intraocular pressure (b) *Constriction of the pupil* (myosis). The pupil of the affected side is smaller than its fellow of the opposite side as a result of the unopposed tonic action of the pupillo-constrictor center (see p 1175) (c) *Vasodilatation, higher temperature and absence of sweating* over the affected side of the face. Irritation of the cervical sympathetic tends to cause the opposite effects, namely, widened palpebral fissure, exophthalmos, dilated pupil (mydriasis) and excessive secretion of sweat on the affected side. In the dog stimulation of the cervical sympathetic causes marked protrusion of the eyeball. The movement is due to the contraction of the circularly disposed smooth muscle fibers in the fascia bulbi. A lesion of the brain-stem or spinal cord may interrupt the central course of the sympathetic fibers to the head, with the production of Horner's syndrome

The relations of the autonomic system to Raynaud's disease (p 297), asthma (p 428), megacolon (p 591) and spastic states of the skeletal muscles (p 963) have been considered in other sections

on to the inner surfaces of the eyelids. The conjunctival surfaces are lubricated and kept clean by a film of fluid secreted by the *lacrymal gland* (fig 73 2). The lacrymal gland is about the size and shape of a shelled almond. It lies under the shelter of the bone forming the upper and outer part of the orbit (i.e., the zygomatic process of the frontal bone). It is of the racemose type, somewhat resembling in structure a serous salivary gland, its secretion—the tears—is delivered through a number of fine ducts into the conjunctival fornix. The secretion is a clear watery, slightly hypertonic fluid having, according to Ridley, the following composition:

	per cent
Water	98.2
Total solids	1.8
Ash	1.05
Total N	0.158
Non-protein N	0.051
Urea	0.03
Protein (albumin and globulin)	0.669
Sugar	0.65
Chlorides (as NaCl)	0.658
Sodium as Na ₂ O	0.60
Potassium as K ₂ O	0.14
Ammonia	0.005

A sample of tears collected from the conjunctival surface contains traces of mucus secreted by the conjunctiva itself. The tears also contain the bacteriolytic enzyme lysozyme (ch 37).

Several small accessory lacrymal glands are situated in the conjunctival fornices; their secretion suffices for lubrication and cleansing under ordinary circumstances. The main glands are called into play only upon special occasions, e.g., irritation of the conjunctiva, as a result of pain, certain emotional states, such as grief, disappointment, anger, etc., and during the acts of yawning and coughing.

The winking movements of the lids spread the tears over the conjunctival surfaces; the fluid is directed into the *lacrymal lake*—a small triangular area lying in the angle bounded by the innermost portions of the lids. The center of the lacrymal lake is occupied by a small pink structure the lacrymal caruncle, composed of modified skin, and containing sebaceous glands and a few slender hairs. The tears are drained from the lacrymal lake by two small tubes—the *lacrymal ducts*. The minute orifices of the latter—the *puncta lacrymalia*—may be seen one on the margin of each lid. The lacrymal ducts lead into the upper part of the *nasolacrymal duct*, this opens into the inferior meatus of the nose, its upper blind end is termed the *lacrymal sac*.

The drainage of tears into the nose does not depend merely upon gravity. Fluid enters and passes along the lacrymal ducts by capillary attraction aided by aspiration caused by contraction of a part of the orbicularis oculi muscle which is inserted into the lacrymal sac

(pars lacrymalis muscle). When the lids close, contraction of this muscle causes dilatation of the upper part of the sac and compression of its lower portion. Tears are thus aspirated into the sac, and any which have collected in its lower part are forced down the nasolacrymal duct towards its opening into the inferior meatus of the nose. As the lids open the muscle relaxes. The upper part of the sac then collapses and forces fluid into the lower part which at the same time is released from compression. Thus, the act of blinking exerts a suction-force-pump action in removing the tears from the lacrymal lake and emptying them into the nasal cavity.

The secretory fibers to the lacrymal gland are derived from the parasympathetic. They arise from the superior salivatory nucleus, or, according to some, from a separate group of cells (*lacrymal nucleus*) in close relation to the latter. The fibers leave the brain in the nervus intermedius of Wrisberg, the sensory root of the facial. They pass to the geniculate ganglion which they leave in the great superficial petrosal nerve (see fig 66 10, p 1000). This nerve joins the deep petrosal nerve to form the nerve of the pterygoid canal (Vidian nerve). The fibers are conveyed in the latter nerve to the sphenopalatine ganglion and thence into the zygomatic branch of the maxillary nerve. A branch of the zygomatic nerve (zygomaticotemporal) anastomoses with the lacrymal nerve—a branch of the ophthalmic. The lacrymal nerve thus receives the parasympathetic fibers and delivers them to the lacrymal gland, it also carries sensory fibers to the gland.

The sympathetic fibers are derived from the cervical sympathetic; they pass into the carotid plexus and travel in the deep petrosal nerve to the great superficial petrosal nerve. They accompany the parasympathetic fibers to the gland. The sympathetic is probably purely vasomotor in function; it does not appear to furnish secretory fibers to the lacrymal gland.

Lacrymation is induced reflexly by stimulation of nerve endings of the cornea or conjunctiva (ophthalmic division of the 5th nerve). The reflex is annulled by anesthetization of the surface of the eye, by section of the sensory nerves or of the great superficial petrosal nerve, or by blockage of the sphenopalatine ganglion. Emotional lacrymation is not affected by local anesthetization nor by section of the ophthalmic division of the 5th nerve, but is abolished by section of the great superficial petrosal nerve or by blockage of the sphenopalatine ganglion. Excessive lacrymation may follow a lesion of the facial fibers central to the facial ganglion. Defective or complete absence of a lacrymatory response, either psychic or reflex is seen, though very rarely, as a congenital anomaly.

THE TUNICS OF THE EYEBALL

The wall of the eye is composed of three concentric layers or tunics: (1) An outer or fibrous tunic—the *sclera and cornea*, (2) a middle vascular

tunic—the *choroid*, *ciliary body* and *iris*, and (3) a nervous tunic—the *retina* (see fig 73 1)

THE OUTER OR FIBROUS TUNIC. The posterior $\frac{2}{3}$ of this coat is opaque and is called the *sclera*, it is composed of white fibrous tissue and fine elastic fibers. Its anterior $\frac{1}{3}$ is perfectly transparent and is called the *cornea*. The sclera appears in front as the so-called “white of the eye”. The point where it joins the cornea—the *sclerocorneal junction*—is marked by a faint groove (see p 1164). The sclera where it is pierced by the optic nerve is reduced to a thin membrane containing perforations for the transmission of the retinal vessels and the bundles of nerve fibers. This part of the outer tunic, known as the *lamina cribrosa*, is the weakest part of the wall of the globe and is the first to yield to a persistently high intra-ocular pressure (see Glaucoma, p 1165),

At the circumference of the cornea the posterior elastic lamina breaks up into fibers which are continued into the pectinate ligament (ch 75). The cornea is devoid of blood vessels, it receives nourishment from lymph derived from vessels at its margin, and which percolates through the spaces between its cells. It is supplied around its circumference by a rich plexus of pain fibers. Fine non medullated filaments derived from this plexus pass through the posterior elastic lamina and form a second plexus in the substantia propria (*stroma plexus*). From the stroma plexus fibers proceed outwards through the anterior elastic lamina where they form a subepithelial plexus, nerve filaments can be traced from the latter to the epithelial cells. The pain fibers have a very low threshold being aroused by very mild forms of stimulation. This has led to the general belief that the cornea is devoid of touch receptors and that

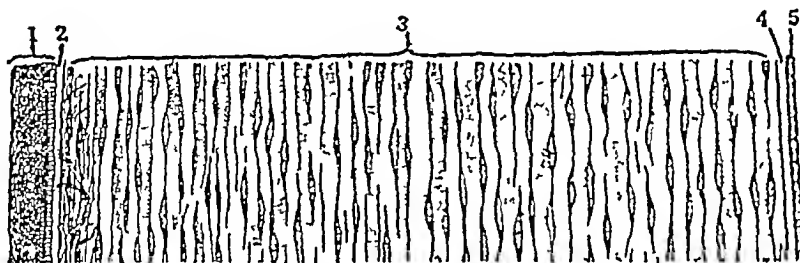


FIG 73.3 Vertical section of human (lying horizontally), 1, corneal epithelium, 2, anterior elastic lamina, 3, substantia propria, 4, posterior elastic lamina, 5, endothelium of the anterior chamber (After Waldeyer, modified)

producing the so-called “cupping” of the optic disc. The cornea is convex anteriorly, being the small segment of a sphere having a radius of about 7.7 mm.² It is almost circular in circumference, measuring 11 mm and 12 mm respectively in its vertical and horizontal meridians. It is from 0.5 mm to 1 mm in thickness. The cornea is composed of five layers in the following order from before backwards, (a) the corneal epithelium, (b) the anterior elastic lamina of Bowman, (c) the substantia propria, (d) the posterior elastic lamina of Descemet, and (e) a layer of endothelial cells (fig 73.3).

The *corneal epithelium* is continuous with that of the conjunctiva, it consists of several strata of cells of different sizes and shapes. Columnar cells compose the deepest layer, this is overlaid by two or three layers of polyhedral cells. The cells of the superficial three or four layers are of the squamous type. The *substantia propria* is a tough transparent membrane consisting of a number of flattened lamellae composed of bundles of modified connective tissue fibers continuous with those of the sclera. The *anterior elastic lamina of Bowman* and the *posterior elastic lamina of Descemet* bound the corresponding aspects of the substantia propria.

² The peripheral zone is somewhat flattened as compared with the central portion, the former has a radius of about 6.8 mm.

stimuli which give a sensation of touch when applied to the skin are painful if applied to the cornea. It is claimed, however, that certain weak and innocuous stimulating agents, such as a jet of fluid impinging upon the cornea, arouse a sensation of touch alone.

THE MIDDLE OR VASCULAR LAYER consists from behind forwards of the *choroid*, *ciliary body* and the *iris*. The choroid is composed of a rich capillary plexus and the numerous small arteries and veins leading to and from it. It is dark brown in color, due to the presence of pigment cells, and forms the middle layer of the posterior $\frac{2}{3}$ of the globe, it terminates anteriorly at the level of the ora serrata of the retina. The ciliary body and iris are described elsewhere.

THE NERVOUS TUNIC OR RETINA. The retina is composed of seven layers of *nervous elements*, an outermost layer of *pigment cells* and two *supporting membranes*. The following is a list of the ten retinal layers from within outwards (see also fig 73.4)

- 1 Internal limiting membrane
- 2 Layer of optic nerve fibers (stratum opticum)
- 3 Layer of ganglion cells
- 4 Inner plexiform or reticular layer
- 5 Inner nuclear layer (bipolar cell layer)
- 6 Outer plexiform or reticular layer
- 7 Outer nuclear layer

- 8 *External limiting membrane*
- 9 *Layer of rods and cones*
- 10 *Pigment layer (stratum pigmenti)*

The *layer of optic nerve fibers* is composed mainly of the naked axons of the ganglion cells composing the subjacent layer. They make a sharp turn a short distance from their origins and converging towards the posterior part of the globe form the optic nerve.

The layer of ganglion cells There are several (five at least) varieties of these cells. Some are large (giant ganglion cells), others quite small (midget ganglion cells). They also differ in shape, oval, pyriform and multipolar forms can be distinguished. The axons of the ganglion cells compose the first layer mentioned.

mop bipolar makes connections mainly with groups of rods and the flat bipolar with both rods and cones (b) *Horizontal cells* These are small flattened stellate or larger irregularly shaped elements whose axons run horizontally. They are purely associative in function, their short dendritic processes forming contact with a cone fiber, and their axons to a more distant group of rods and cones (c) "*Amacrine*" cells Many of these appear to possess no axons. Their dendritic processes for the most part ramify in the inner plexiform layer (*diffuse* type). The dendrites of others do not leave the inner nuclear layer (*stratified* type). The axons of some can be seen ramifying in the outer plexiform layer. The polarity of the amacrine cells appears to be in the

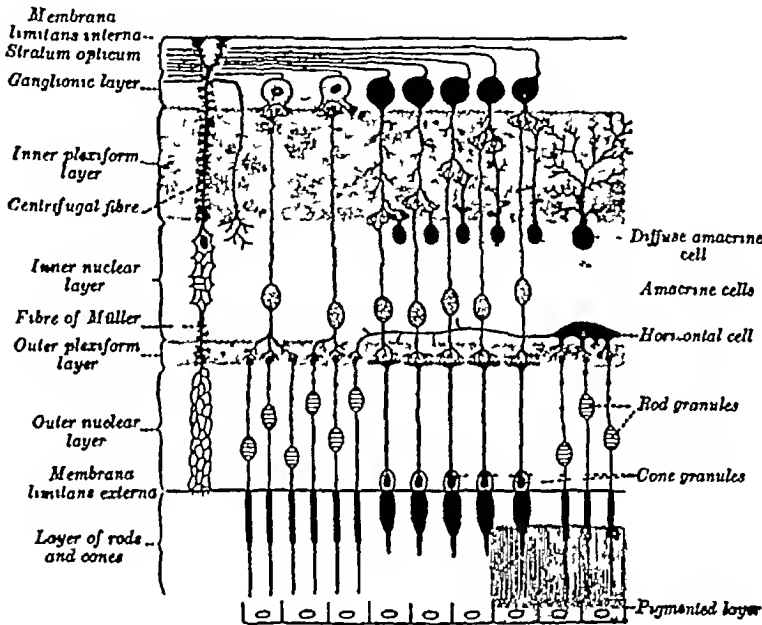


FIG 73 4 Layers of retina (After Cajal)

above, their dendrites ramify in the inner plexiform layer where they form synaptic connections with the bipolar cells (see below). Each giant ganglion cell may receive impulses from both rods and cones through the mop and flat bipolar cells of the inner nuclear layer. But the "midget" ganglion cells connect each with a single "midget" bipolar, which, in turn, provides a private line from individual cones (see fig 73 5).

The *inner plexiform layer* is made up mainly of the dendrites belonging to the ganglion cells and to the cells of the inner nuclear layer. There are, in addition, a few small horizontal nerve cells possessing short branching processes.

The *inner nuclear layer* is constituted of three types of cell (a) *Bipolar cells* These possess oval bodies and two long slender processes which arise from opposite poles. Three main types are described by Polyak, *midget*, *flat* and *mop* bipolars. The midget bipolars are plentiful in the central part of the fovea, and as mentioned above convey impulses from single cones. The

centrifugal direction, and, therefore, opposite to that of the bipolar and ganglion cells, which convey impulses centripetally.

The *outer plexiform layer* is made up of the ramifications of the processes of the bipolar and horizontal cells of the inner nuclear layer and the terminations of the rod fibers and cone fibers.

The *outer nuclear layer* is composed of the cone granules and fibers and the rod granules and fibers.

The nervous elements of the various layers are supported by a framework of fibers of neuroglial character (sustentacular fibers of Mueller). They extend from the inner aspect of the stratum opticum where they constitute the *internal limiting membrane* to the bases of the rods and cones where they break up into a fine feltwork of fibers to form the *external limiting membrane*.

The *layer of rods and cones* The *rod* and *cone* cells (fig 73 5) are the visual receptors. Both these elements lie with their long axes perpendicular to the retinal surface. The rod cell (40 μ to 60 μ long) consists of two

well-defined portions, a cylindrical outer segment—the *rod*—which extends from the external limiting membrane to the pigment layer, and an inner segment—the *rod fiber*. The latter is a long slender filament swollen at a variable distance along its course by the cell nucleus—the so-called *rod granule*. The inner end of the rod fiber shows a slight enlargement called the *end button* or *spherule* which makes connection with the arborizations of the bipolar and horizontal cells in the outer plexiform layer. The rod itself shows two segments of about equal length, but the outer one is only about half the thickness of the inner and is marked by transverse striae. The two parts of the rod differ

The number of receptors in the human retina according to most recent estimates is about 115,000,000 rods and 6,500,000 cones. In the primate fovea each cone cell has its own private pathway to the brain, via a midjet bipolar and a midjet ganglion cell with its optic nerve fiber, whereas, in the extrafoveal region a single fiber in the optic nerve carries impulses from a number of rods or of rods and cones. There appear to be three main pathways from the retina, (a) a pure cone pathway via midjet bipolar and midjet ganglion cell, (b) a primary rod path via the mop bipolar and the giant ganglion cells, and (c) a common cone rod path via the mop bipolar and a giant ganglion cell.

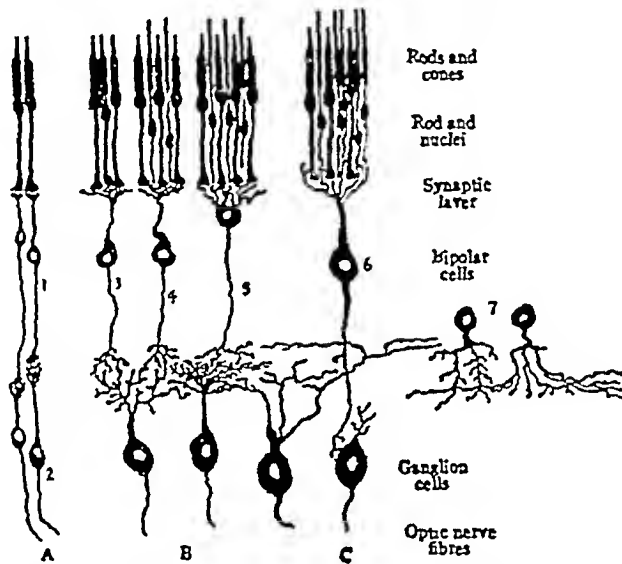


FIG 73.5 Chief neural connections of the rods and cones (From Polyak, modified) A, pure cone pathway, mixed rod and cone pathway, C, rod pathway, mainly 1, midjet bipolar cell, 2, midjet ganglion cell, 3, flat bipolar supplying cones, 4, flat bipolar supplying rods and cones, 5, brush bipolar supplying rods and cones, 6, mop bipolar cell, 7, amacrine cells. The brush bipolar is a variant of the flat bipolar.

chemically, each reacting in its own way to certain dyes, the outer thinner portion is composed of a myelin like material, the inner, of protoplasm. The outer-segment breaks up after death along the striae just mentioned into tiny transverse plates, this part of the rod alone contains a reddish pigment—the *visual purple* or *rhodopsin*.

The cone cell consists of a pyramidal portion—the *cone*—situated on the outer side of the external limiting membrane with its pointed end directed towards the pigment layer, and an inner segment—the *cone fiber*—which varies in length and thickness according to the part of the retina in which it is situated. The cone fibers contain the nuclei, these are the *cone granules* already mentioned with the rod granules as being the chief elements of the outer nuclear layer. The cone cells vary from 28 to 85 microns in length in different parts of the retina, and from 2.5 to 7.5 microns in thickness.

It is evident that light in order to reach the rod and cone layer must, except at the fovea, pass through all the other nervous layers. The retina has been compared to a transparent carpet laid upside down (Walls). The pile then corresponds to the visual receptors and the jute backing to the maze of nerve fibers which support the visual and integrate the impulses initiated in them.

The *pigment layer* consists of a single row of epithelial cells containing mobile rod shaped granules of a dark brown melanin like pigment called *fuchsin*. The outer surfaces of the pigment cells are hexagonal in outline (fig 73.6) and are firmly attached to the choroid. From their inner portions fine processes arise which are insinuated between the rods and cones. In darkness or dim light the pigment granules are aggregated in the inner half or so of the cell body but in certain species, e.g., the frog and fishes—not in mammals—(p 1130), illumination of the eye causes them to move into the processes between the visual cells. The pigment serves

like the black print on the inside of a camera, to absorb light which otherwise would be diffused and cause blurring of the retinal image. In albinos the hexagonal cells are free from pigment.

DEVELOPMENT OF THE RETINA AND DESCRIPTION OF ITS GROSS APPEARANCE Knowledge of the origin of the retina is an important step towards an understanding of its functions. It is developed from a hollow outgrowth of the rudimentary fore brain called the *optic vesicle*. The distal portion of this diverticulum expands, while the proximal portion remains narrow and is termed the *optic stalk*. As development proceeds the outer wall of the optic vesicle collapses and becomes invaginated, finally, it meets the inner wall and joins with it, thus a two layered cup shaped structure—the *optic cup*—is formed. The outer layer consists of a single row of epithelial cells which acquire pigment, and is seen in the adult retina as the *stratum pigmenti*. The inner layer formed from the invaginated wall becomes thickened, and from it are developed the nervous layers of the retina. After death the nervous layers (retina proper) can be readily stripped up from the pigment cells which remain attached to the choroid. The membrane freshly detached from the pigment layer appears reddish (due to the presence of visual purple) and, transparent when shielded from strong light. Upon exposure to bright light, however, it rapidly becomes colorless (bleached) and clouded. The line of union between the two walls of the optic cup is always a weak part in the adult retina, and in the living eye detachment of the retina from the pigment layer is not an uncommon accident.

The retina proper extends from the margins of the optic papilla (see below) to just behind the ciliary body. At this point it ends abruptly in a dentated border—the *ora serrata* (see fig 75 29, p 1152). Its thickness diminishes progressively from the optic papilla (where it measures about 0.4 mm) to the dentate border where it is only 0.09 mm thick. The pigment layer is continued forwards from the ora serrata over the deep surface of ciliary body (*pars ciliaris retinae*) and iris (*pars iridica retinae*). Two retinal areas require special mention—the *optic disc* and the *macula lutea* with the *fovea centralis*.

The *optic disc* is situated about 3 mm to the nasal side of and a little above the posterior pole of the eyeball, it has a diameter of about 1.5 mm in man. As viewed in the human eye by means of the ophthalmoscope (p 1147) it appears as a pink circular area fading to a creamy white toward the center. It is pierced near its center by the retinal vessels—*arteria centralis retinae* and its accompanying vein (Pl III, opp p 1146). The circumference of the optic disc is elevated to form the *optic papilla*. The central depressed part is known as the *physiological cup* or the *excavation of the optic nerve*. The vessels climb up the inside of the cup to reach the retina. All layers of the retina except the nerve fiber layer are absent from the optic disc. It is therefore

totally insensitive to light and is known as the *blind spot* of the retina. The reader is referred to figure 76 2 for a demonstration of the blind spot in his own eye.

The macula lutea A small diffuse yellow area called the *macula lutea* (yellow spot) is seen in the retina a little to the temporal side of the posterior pole of the globe, i.e., about 3.5 mm from the outer edge of the optic disc. Its color is due to the presence of a yellow pigment which, like the visual purple, is bleached by light, though much less readily. The central zone of the macula lutea is depressed and is referred to as the *fovea centralis*. This is the region of the retina concerned

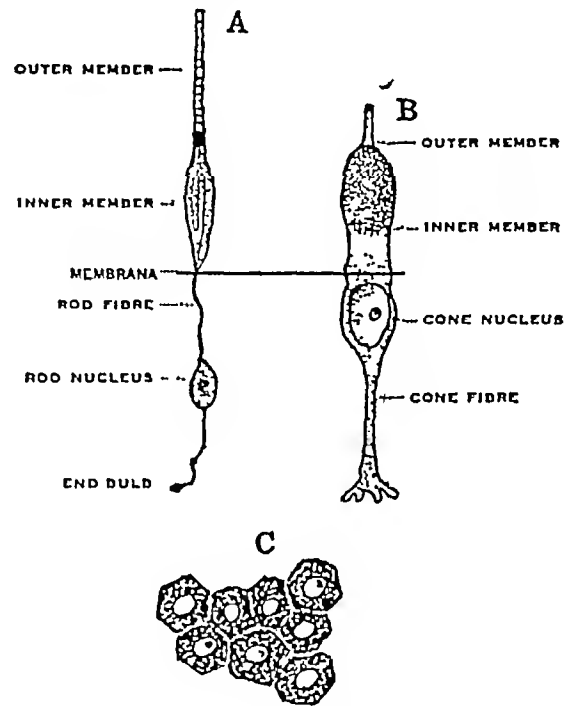


FIG 73 6 A, rod, B, cone, C, pigment cells viewed from the surface

with acute vision, in man it has a diameter of about 0.44 mm (Wolfrum). The depression is due to the extreme thinning of the retina proper at this point, brought about by the disappearance of the nerve fiber, ganglion cell, inner plexiform, inner nuclear and outer plexiform layers. The elements which compose these layers are swept peripherally, leaving only the cone cells which are here exceptionally long (85 μ) and slender (fig 73 7). At the bottom or center of the fovea the thickness of the retina is only 0.1 mm. Moreover, the cone fibers of this region are elongated and, instead of passing perpendicularly inwards, incline obliquely to connect with the bipolar cells around the margins of the fovea. In this way a layer is formed which, though it corresponds to the outer nuclear layer, differs in appearance from that of any other part of the retina. It is called the *outer fiber layer of Henle*. In this region it alone overlies the cones. The pigment layer is exceptionally well developed at the fovea.

The human fovea is rod free. In the remainder of the retina the cones diminish progressively in number from the foveal margin to the periphery, while the rods increase proportionately. The macula immediately surrounding the fovea is nearly but not quite rod free. In some animals, e.g., the rat and certain night-flying birds, the retina throughout contains only rods, in others, especially day birds, rods are almost entirely absent.

It must never be forgotten that the retina is an outgrowth of the brain and retains the structural and functional characteristics of nervous tissue. The neurons composing its several layers are connected through synaptic junctions, and such phenomena as spatial summation, latency, inhibition, convergence and occlusion (see chapter 64) have been demonstrated in the retina.

THE OCULAR CIRCULATION Thirty three separate arteries enter the eye ball. The *retina* is supplied by the *arteria centralis retinae*, a branch of the ophthalmic

the tissue of the iris to the pupillary margin, here they join to form a smaller arterial ring (*circulus arteriosus minor*). The short posterior ciliary arteries perforate the sclera around the optic nerve and supply the choroid and ciliary process. The anterior ciliary arteries and their companion veins pierce the globe a little behind the sclerocorneal junction. The blood is returned from the choroid by a system of veins in the outer choroidal layer. From their whorl-like arrangement they are termed the *vortex veins* (*venae vorticales*). The smaller and medium sized vessels of this system become confluent to form four trunks which penetrate the sclera and appear on the surface of the globe equidistant from one another just behind its equator. These vessels and the anterior ciliary veins drain into the ophthalmic veins. The central vein of the retina empties into the cavernous sinus either directly or through one of the ophthalmic veins (p. 1151).

In man the pressure in the central artery of the retina is from 70 to 85 mm. Hg systolic and from 40 to 50 mm

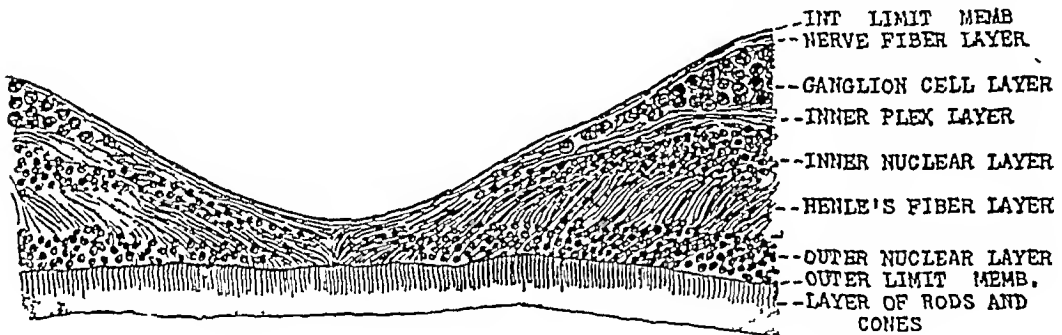


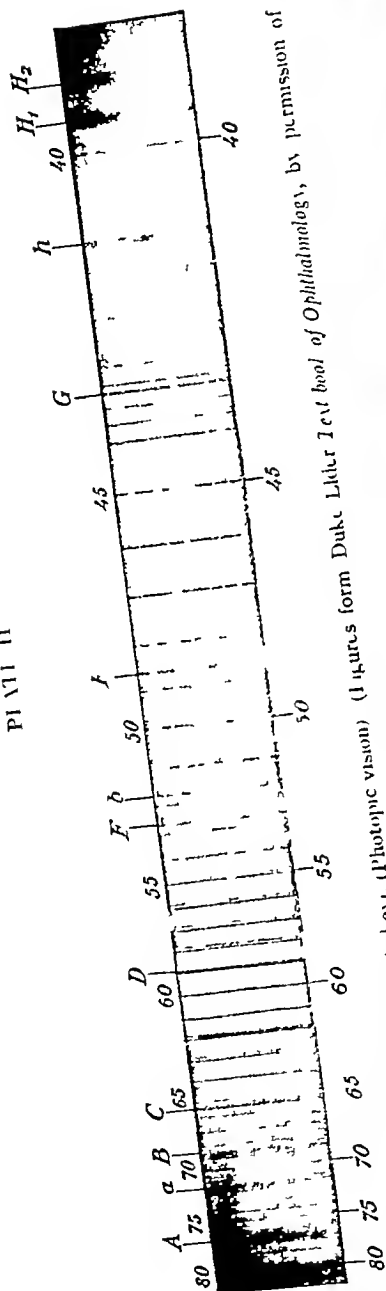
FIG 737 Section through the center of the fovea centralis (From Wolff, after Eisler)

artery. The central artery with its companion vein pierces the optic nerve about 1.25 cm. behind the eyeball and bending sharply runs forwards in the center of the nerve. Perforating the lamina cribrosa it appears inside the eyeball at the center of the optic disc. It immediately divides into two main branches which redivide to form a vascular network, the finer channels ending in a capillary plexus which extends outwards as far as the inner nuclear layer. The fovea itself is devoid of vessels. The *choroid*, *iris* and *ciliary body* are supplied by a separate system of vessels derived from the long and short posterior ciliary arteries, branches of the ophthalmic, and from the anterior ciliary arteries which are twigs of the lacrymal branch of the ophthalmic. The two long posterior ciliary arteries pierce the sclera a short distance from the optic nerve and, running forwards on either side of the globe between the choroid and the sclera, anastomose with branches of the short posterior ciliary arteries and of the anterior ciliary artery to form a vascular ring (*circulus arteriosus major*) which encircles the periphery of the iris. Branches pass from this vascular circle along converging lines through

diastolic. But owing to the presence of the intra-ocular fluid and the resistant nature of the sclerotic coat, pulsation of the retinal artery as observed by ophthalmoscopic examination is slight. The pressure in the retinal artery of the intact globe may be determined by the method of Baillart which consists in observing the vessel with the ophthalmoscope while a measured pressure is made upon the globe. Maximal pulsation is taken as indicating the diastolic level and the disappearance of pulsation as an index of the systolic pressure. There are several fallacies in this or any other indirect method, the results being far from reliable. The pressure in the central vein of the retina is around 25 mm. Hg. A venous pulse is also observed, it is attributed to the transmission of the impulse from the artery through the intra-ocular fluid to the veins. That is to say, with each expansion of the artery the veins for a short distance proximal to where they leave the orbit are compressed and an extra quantity of blood is ejected from the eyeball.

The lens and ciliary body are described in chapter 75, the visual pathway and ocular muscles in chapter 76.

PLATE II



Spectrum as seen by light adapted eye (Photopic vision) (figures from Duke Elder Text book of Ophthalmology, by permission of Henry Kimpton, London)



Spectrum as seen by the dark adapted eye (Scotopic vision)

THE DUPLICITY THEORY OF RETINAL FUNCTION

The theory that the retina is a duplex organ was first proposed by Max Schultz in 1866. There is today abundant evidence for Schultz's doctrine, which is now referred to as the *Duplicity Theory*, that the rods function in dim light, below an illumination of 0.01 foot candle (*scotopic vision*), and do not register sensations of color, whereas the cones are stimulated only by higher levels of light intensity (*photopic vision*), and are responsible for the perception of color. The fovea, which contains but few if any rods, is the region of the most acute vision. The surrounding (peripheral or extrafoveal) retina is composed of both rods and cones, the proportion of the former type of element increasing, that of the latter diminishing towards the periphery where cones are entirely absent.

The sensitivity of the rods is dependent upon a photochemical substance (see visual purple). Though rods "take over" from the cones as the intensity of the light is reduced below the threshold of the latter, they do not cease to function when, at a higher level of illumination the cones become active. Within a certain range both types of receptor respond, but as the intensity of the light is further increased the rods cease to react.

At the risk of some repetition in other sections the evidence for the duplex nature of retinal function will be briefly reviewed.

(1) *The Purkinje shift* When the spectrum is viewed in a bright light (i.e., by the light adapted eye, p. 1125) it is seen as a series of colors with the maximum brightness (luminosity) in the yellow (sodium D-line). If the illumination of the spectrum is reduced and the eye dark adapted, the region of maximum luminosity will be found to have shifted nearer to the violet end. The red portions gradually become darker, the blues correspondingly brighter, and with further reduction in the intensity of the light a point is reached where the spectrum is colorless, the maximum luminosity is now in the region of the E line (see Plate II and fig. 73.8).

These facts explain the familiar visual experience that red and blue objects though equally bright in daylight, show different degrees of brightness in twilight. As night falls the red geraniums in the garden become darker and darker, and finally appear black, and, therefore all but invisible, while the blue delphinium can still be seen as gray patches.

This shift of the spectral region of maximal luminosity with a change in the intensity of illumination is taken as strong evidence for the existence of two types of light sensitive elements in the retina, one type—the cones—being responsive at intensities above 0.01 foot candle, the other type—the rods—reacting to the lower levels

of illumination. The Purkinje phenomenon cannot be demonstrated at the fovea, where rods are absent or very few in number.

The Purkinje phenomenon is shown graphically by means of curves of retinal sensitivity to the different wave-lengths, in bright and dim light, respectively. The curve obtained in bright light is called the *photopic luminosity or visibility curve*, that obtained at low illuminations after dark adaptation, the fovea being unresponsive (blind), is known as the *scotopic luminosity or visibility curve*. In obtaining the data for the curves, monochromatic lights of the different wave-lengths are

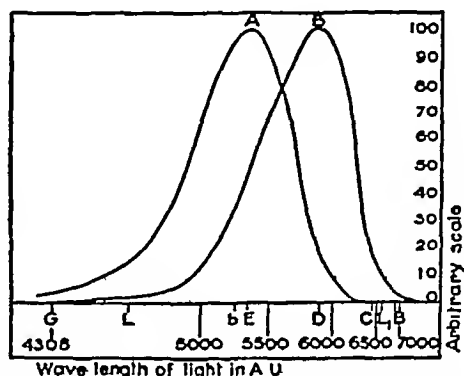


FIG 73.8 Normal scotopic (A) and photopic (B) luminosity curves (After Abney and Festing)

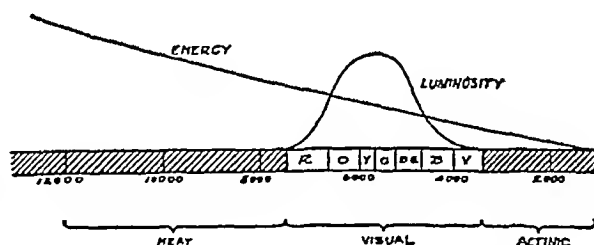


FIG 73.9 Curves showing the relative energy and luminosity of the different regions of the spectrum (Starling, *Principles of Human Physiology*, 3rd Edition, Lea & Febiger)

employed, and, in the case of the scotopic curve, threshold intensities are determined. It is not feasible to obtain a photopic curve in the same way. In this case, a light of a standard brightness is shown in one half of a photometer field and the other half of the field filled successively with monochromatic lights of the various wave-lengths. The intensity of each of the latter lights is altered by the observer until a match in brightness is made. In the construction of either curve the reciprocal of the energy required to make a match, or for a threshold stimulus in the case of the scotopic curve, is plotted against the wave-length. Thus, the sensitivity of the retina is related inversely to the quantity of energy employed—the less the energy required the greater the sensitivity of the retina.

The greater sensitivity of the retina to some wave-lengths than to others is not simply a matter of the total

energy content. A glance at figure 73 9 will show that the distribution of energy in the spectrum is not related to the luminosity curve. In figure 73 10 the luminosity curves are corrected to an equal-energy spectrum. It will be observed that the maximum luminosity is at a wave

length of 560 $m\mu$, it will be seen from the curves that the former has a stimulating power in arbitrary units of nearly 100 at low intensities of illumination (rods), and only about 70 in bright light (cones). Yellow light has a rod stimulating power of about 40

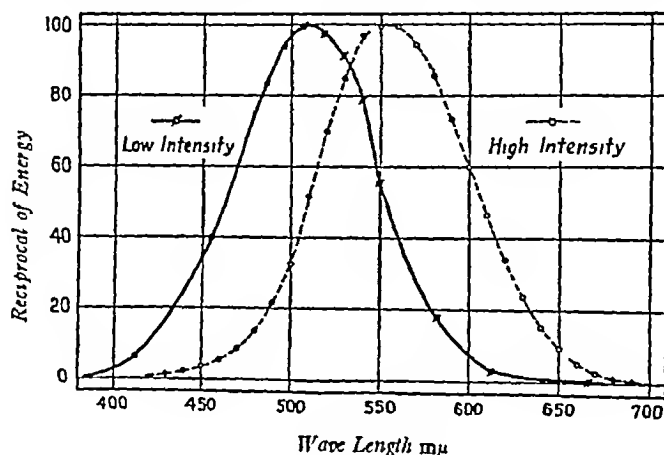


FIG 73 10 Visibility or luminosity curves for scotopic and photopic vision (After Hecht.)

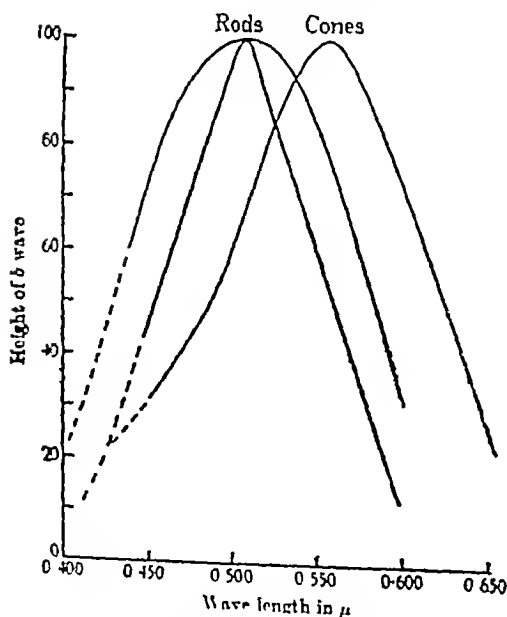


FIG 73 11 Rod and cone curves for size of b wave of electroretinogram plotted against wave length (frog) (From Granit and associates)

length of 560 $m\mu$ for high light intensities, and at 510 $m\mu$ for low. Thus, the stimulating power is represented as a function solely of the wave length. That is to say, if an equal energy spectrum were thrown upon a screen and a comparison made between the stimulating power of green, having a wave length of 510 $m\mu$, and yellow

units as compared with nearly 100 units of cone stimulating power.

(2) Using the electroretinogram (ch 74) as the criterion of retinal response Graham and Riggs found that a Purkinje effect could not be demonstrated in the white rat. The wave length of maximum effectiveness was the same in both the light and dark adapted eye, namely, 510 $m\mu$, which, as just stated, is that for the dark adapted eye of man. The significant point is that the retina of the white rat contains only rods. In such forms as the frog and fish which possess both rods and cones the Purkinje shift can be demonstrated by electrophysiological means (see figure 73 11). On the other hand, the retinas of day birds such as the domestic hen and the pigeon, which contain cones but very few rods, show a spectral sensitivity corresponding to that of the light adapted human eye, whereas night birds¹ such as the owl, and other mammals as well as the rat, e.g., the cat, dog, guinea pig and rabbit, whose retinas contain mainly or only rods, show a spectral sensitivity similar to that of the human retina in dim light.

Potential changes recorded from the human eye caused by the illumination of the retina with red and

¹ The uncanny faculty of some night birds, notably the owl, to see in apparently absolute darkness has suggested that their retinas are stimulated by infra red radiations to which the eyes of mammals are insensitive. Thus, it has been supposed that the bird's prey, e.g., a mouse, having a temperature above that of its surroundings, emits infra red rays which the avian eyes can detect. Careful investigation of this question has proved, however, that the sensitivity of the owl's retina does not extend beyond the visible spectrum.

with blue light give further evidence for photopic and scotopic mechanisms (Adrian) With red light, which stimulates mainly the cones, the record shows a brief diphasic change which is increased only slightly by dark adaption Blue light, which stimulates the rods predominately, results in a slower monophasic response This response is increased greatly by dark adaption

(3) The sensitivity of the human retina beyond the fovea is enormously increased by dark adaptation, whereas the fovea shows relatively little change The color sense is completely lacking from the most peripheral part of the retina which is cone-free

(4) *Night blindness (nyctalopia)* and *day blindness (hemeralopia)* In night blindness, which is usually due to vitamin A deficiency but may be hereditary, vision is defective in dim lights, whereas daylight vision is normal and the spectral sensitivity of the fovea is that of the normal eye The visual purple of the rods supposedly is lacking or regenerates very slowly in the dark after exposure to light In day blindness there are lack of color sense and photophobia These facts accord with the belief in the independent functions of the rods and cones

(5) *Visual acuity* (p 1118) and the *discrimination of light intensity* (p 1121) Investigations of these faculties and of the phenomenon of *flicker* (p 1122) furnish additional evidence of the dual nature of retinal function

(6) The close correspondence between the scotopic luminosity curve and the absorption curve of visual purple (p 1127) is further evidence for the functioning of the rods in twilight.

(7) *The perception of movement* is more acute in the peripheral retina than at the fovea

The differences between the functions of the rods and cones are not quite as definite and clear-cut as the foregoing experimental results would seem to imply There are certain observations which conflict with the generalizations of the classical view, some revision in our ideas is required in the light of recent work Snakes, for example, which have a pure cone retina do not appear to be blind in the dark (in India they invade houses usually at night) and many mammals whose retinas contain cones are apparently without color sense The retina of the guinea-pig which is of the pure cone type, shows as far as dark adaptation is concerned, exactly the same reactions (e.g., Purkinje shift) as the mixed cone-rod retina of the cat (Granit) The conclusion seems inevitable, and it is now generally believed that there are two types of rods which though indistinguishable histologically, differ in function Rods of one type behave like the cones and have been called "day rods", the other type behaves like the ordinary classical rod in so far as

its possessing visual purple, and its reactions in dark adaptation are concerned This latter type of rod is believed to be especially sensitive to the blue end of the spectrum, and is absent from the fovea According to Willmer the fovea is blue blind, the cones and the "day rods" being insensitive to this color

THE REACTIONS OF THE RETINA, SUBJECTIVE PHENOMENA

MECHANICAL AND ELECTRICAL STIMULATION
The visual cells—the rods and cones—are highly specialized for the reception of radiations of the visible spectrum and for the conversion of the radiant energy into nerve impulses The retina, in the expressive phrase of Sherrington, is a "glorified heat spot" But, though it is stimulated most effectively by light, which is therefore the *adequate stimulus*, crude visual sensations can be evoked by mechanical and electrical forms of stimulation Pressure upon the eyeball, for example, causes to appear a luminous ring, known as a *phosphene*, and the flash of light caused by a blow upon the eye is familiar to everyone A phosphene also results from electrical stimulation of the eyeball⁴

As in the case of nerve or muscle, chronaxie is a measure of retinal excitability Verryp and Bourguignon and associates have investigated the chronaxie of the human retina, employing the phosphene response to electrical stimulation There are apparently two chronaxies, one due to the rods, the other to the cones With the electrodes nearer the fovea (cones) the value was from 2.1 to 2.8 milliseconds, when nearer the periphery of the retina 1.0 to 1.9 milliseconds The chronaxies are the same whether the eye is light or dark adapted (p 1124)

STIMULATION OF THE RETINA BY LIGHT

The wave lengths (λ) of radiant energy which stimulate the retina (visible spectrum) range from around 400 μ to 780 μ (4000 to 7800 Angstrom units⁵) This is only about one octave of the entire energy spectrum The last figure just given evidently represents the limit of sensitivity of the retina to the longer wave lengths, for the media of the eye will transmit rays much longer than this

⁴ The three primary color sensations (p 1134) can be elicited by graded pressures upon the globe With a pressure just sufficient to be decidedly uncomfortable violet dots appear which persist for a time after the pressure has been released Stronger pressure causes green and still firmer pressure red figures to be seen

⁵ μ m or millimicron = $1/1,000,000$ mm, an Angstrom unit (A.U) = $1/10,000,000$ mm

On the other hand, the absolute limit of transmissibility of the shorter wave lengths is around 300 $m\mu$, but this figure applies only if the illumination is intense and even then few under 400 $m\mu$ reach the retina. The failure of the retina to respond to rays of waves of shorter length than this is apparently because the ocular media are opaque to them.⁶

The cornea transmits rays from λ 295 $m\mu$ to λ 2,500, but above 1,800 $m\mu$ the transmission is slight. Maximum transmission occurs at around 1,000 $m\mu$ and continues high at lower wave lengths until the ultraviolet is reached. Absorption then becomes pronounced and is complete at about 295 $m\mu$. The aqueous humor and vitreous body are somewhat less transparent than the cornea. All rays above 2,700 $m\mu$ are absorbed by the aqueous humor and all above about 1,600 $m\mu$ by the vitreous. The lens absorbs all rays below 300 $m\mu$, or above 2,500, most of the rays longer than 1,300 $m\mu$ are absorbed. Maximal transparency is between 1,200 and 400 $m\mu$. That is, most of the rays above and below this range are absorbed by the lens. Wave lengths between 350 $m\mu$ and 400 $m\mu$ cause fluorescence in the lens. The phenomenon of fluorescence, in general, is attributed to the transference of the energy of the incident radiations to particles of the substance absorbing them. The particles then act as independent light sources, emitting waves which are, for the most part, longer than the original radiation. Thus, the lens converts the harmful shorter waves to longer ones which are permitted to reach the retina. That wave lengths shorter than those which are transmitted by the lens are capable of stimulating the retina is shown by the fact that the visible spectrum is extended towards the blue end after removal of a lens of normal transparency. The lower limit of transparency of the lens rises with age. In the aged it is opaque to rays longer than about 400 $m\mu$. In the early stages of cataract the lower limit is around 450 $m\mu$ (fig. 73 12).

The quantity of light energy⁷ necessary to stimulate the retina is not constant for all wave

lengths of the visible spectrum. In the dark adapted eye, the energy required to stimulate the rods and cause a just perceptible sensation is least for green light with a wave length of about 507 $m\mu$ (see p. 1113) and has been estimated by Hecht at from 2.1 to 5.7×10^{-10} Erg,⁸ measured at the surface of the cornea in the dark adapted eye. This equals from 58 to 148 quanta of light energy. After allowance was made for loss of energy in transmission to the retina through the ocular media and for incomplete absorption by the rods, the value was only from 5 to 14 quanta, or one quantum absorbed per stimulated rod. The extraordinary sensitivity of the retina to light may be expressed in simpler terms by saying that light emitted by a standard candle at a distance of nearly a mile would be visible if the air were perfectly transparent. Such examples indicate that the eye has a sensitivity some 300,000 times greater than that of the most deli-

h is Planck's constant of action, 6.62×10^{-27} erg \times second. The frequency ν is equal to the velocity of light (2.998×10^{10} cm/sec.) c , divided by the wave length λ . The energy content of a quantum of light is, therefore, inversely proportional to the wave-length of the radiation. Thus, a quantum of green light has a greater energy content than one of orange or red. The meter candle is a unit which is in part subjective and in part objective. It is a measure of surface illumination, being defined as the light incident per second per sq. cm. of a surface placed at right angles to the beams from a standard (international) candle 1 meter distant (a standard candle is made of spermaceti, weighs $\frac{1}{2}$ lb., and burns 120 grains of wax per hour with a flame 45 mm. high). Illumination is measured by means of a photometer. If, for example, the illumination of a lamp is to be measured, a photometer is set up at this point, and a standard candle is moved until its light just matches that of the lamp. If the candle's distance in meters is d , then the surface illumination is $1/d^2$. The meter candle is thus a measure of the visual stimulus, not of the sensation itself. The quantity of light reflected from a perfectly diffusing surface (e.g., of magnesium oxide) 10,000 meter candles is termed a lambert (l), a millilambert (ml) is 1/1000 lambert. The lumen is the unit of emitted light (luminous flux). One lumen is the light emitted in a unit of solid angle by a uniform point source of 1 standard candle. A point source is regarded as occupying the center of a sphere, a unit solid angle is the angle subtended at the center of the sphere by an area on its surface equal to the square of the radius. The area of a sphere is $4\pi r^2$, therefore 1 standard candle emits 4π lumens. The intensity of a light source is expressed in lumens, i.e., the luminous flux emitted in any direction per unit of solid angle.

The unit of retinal illumination is called the *photon*, or, preferably, the *Troland*. The amount of light which illuminates the retina depends not only upon the brightness of the object but also upon the size of the pupil. A Troland is, therefore, defined as the illumination of the retina by light from a surface having the brightness of one standard candle per square meter, as seen through a pupil having an area of 1 square millimeter.

⁸ The quantum of green light is 3.84×10^{-13} ergs

⁶ The X rays, of course, penetrate the eye and to such the retina is sensitive. This may be demonstrated by means of metal letters or figures placed in front of a normal eye with closed lids or before any eye in which the retina is normal. A word can be read in this way, the letters when exposed to the rays are clearly recognized as shadows (not images) against a bright ground. The rays are not refracted and the shadows are not inverted upon the retina as are ordinary retinal images, therefore, if they are to appear upright to the subject they must be placed in the reverse position in front of the eye (see p. 1146).

⁷ Radiant energy is expressed in *ergs*. The erg is a purely objective or physical unit, being quite independent of visual sensations. The energy is measured by means of a thermopile, bolometer or radiometer. Light is radiated in elementary units called *quanta*. The energy content of a quantum of light is proportional to the frequency of the radiation, ν , and equal to $h\nu$, where

cate radiometer and about 3000 times greater than that of a rapid photographic film Pirenne has made the interesting calculation that the mechanical energy of a pea falling from a height of 1 inch, if converted to luminous energy would be sufficient to cause a faint visual sensation "to every man that ever lived" The minimum quantity of light energy required to evoke a visual sensation increases progressively towards either end of the spectrum, being several thousand times greater for red and blue than for green light

Anyone who has used a camera will recognize that with respect to the relationship between intensity and duration the retina behaves like a photographic film—the lower the illumination the longer must be the exposure, and vice versa As a matter of fact, the retina obeys the Bunsen-Roscoe law applicable to photosensitive reactions in general (*intensity × time = constant*) In the case of the retina this relationship holds only for periods of less than about $\frac{1}{4}$ of a second Also, the relationship between intensity and area is valid

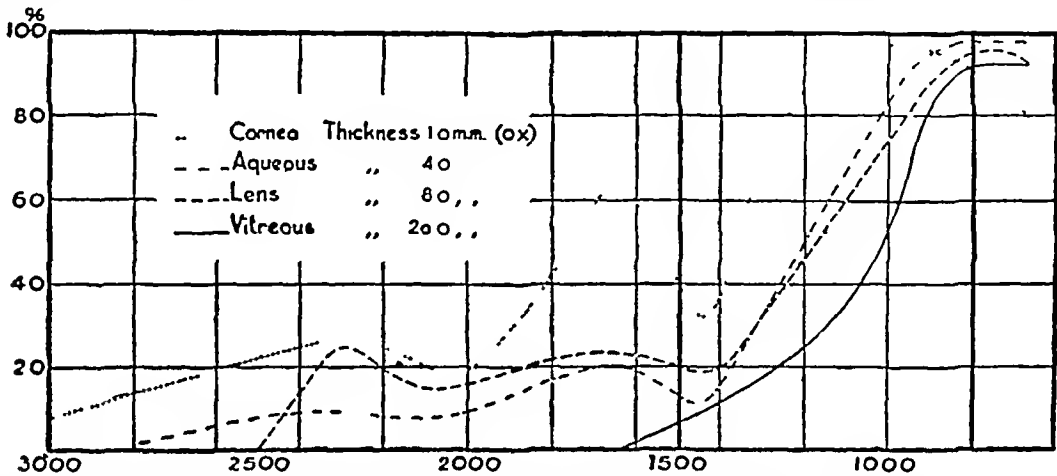


FIG. 73.12 The transparency of the ocular media for the longer wave-lengths of the spectrum (From Duke Elder, after Roggenbau and Wetthauer) Wave-lengths in μ along the abscissae

The quantity of light energy, i.e., the intensity of the light in the physical sense, required to cause a just perceptible sensation—the *intensity threshold*—is therefore a function of the wave length The sensitivity also varies in different regions of the retina with the state of adaptation of the eye and with the illumination of the surrounding field

The intensity threshold is affected within certain limits by two factors, (a) the size of the light source, i.e., of the retinal area stimulated, and (b) the duration of the light stimulus The quantity of light required to stimulate the retina is constant Thus, with respect to (a) if a light source of a certain size and intensity causes a just perceptible visual sensation, reducing either its intensity or its size, renders it invisible This is known as *Ricco's law* and for the fovea is expressed by the following formula

$$\sqrt{\text{threshold intensity}} \times \text{angular}$$

diameter of the retinal area = constant For the peripheral retina the formula is

$$\text{threshold intensity} \times \sqrt{\text{area stimulated}} = \text{constant}$$

only for retinal areas under about 0.8 sq. mm. It would appear from the foregoing consideration that the total quantity of light energy falling upon the retina within a given time is the fundamental factor which determines the retinal response

The smallest area which can be perceived is called the *minimum visible* The angle subtended at the nodal point of the eye by the minimum visible is referred to as the *minimum visible angle* But from what has been said it is quite evident that neither of these can be given any absolute value A light could be reduced in size indefinitely—to a mathematical point—and, provided it were sufficiently intense, would still be visible, a star, for example, subtends an infinitesimal visual angle

THE DIRECTIONAL SENSITIVITY OF THE RETINA
Stiles and Crawford have shown that a pencil of light which enters the eye through the center of the pupil causes a more intense sensation than does one of equal physical brightness passing obliquely through the pupil nearer its circumference This phenomenon is most pronounced at the fovea The smaller effect caused by the oblique beam is not due to greater absorption by the ocular media, but to the direction of the beam in relation

to the position of the cones, a smaller photochemical change being caused by light passing obliquely across the receptors than by that which traverses their lengths

VISUAL ACUITY, THE RESOLVING POWER OF THE EYE The acuteness of vision (or visual acuity) is dependent upon several retinal functions, e.g., the sensitivity to light (intensity threshold), the minimum visible and the ability to recognize the separateness of two closely approximated or parallel lines. The threshold of the latter faculty is commonly referred to as the *minimum separable*⁹ or the *resolution threshold*. Visual acuity is the basis of the *form sense*, by which is meant the power of determining by sight the shape, form, outline and minute detail of our surroundings. Visual acuity is customarily expressed in terms of the minimum separable or, to be more explicit, as the reciprocal of the angle subtended at the nodal point of the eye—the *visual angle*—by the space between two points situated at the minimum distance apart at which their duality can be recog-



FIG 73 13 See text.

nized. For example, if the visual angle is 1/321 minute, then the visual acuity is $(\frac{1}{\frac{1}{321}} =) 0.756$. The average normal eye can resolve two points when the visual angle is 1 minute (60 seconds). The minimum angle that has been reported is 44 seconds (vis. ac. 1.492). As an object is moved away from the eye its visual angle becomes progressively smaller. Consequently, those details of form and structure which subtend an angle of a minute or more at the nearer point and are therefore visible, gradually become imperceptible with increasing distance. In other words, in order to see an object at a distance as clearly as when it is near the eye, it would need to be increased proportionately in size (see fig. 73 13).

In determining the visual acuity, figures such as the broken circle **C** of Landolt or Snellen's prong, **E**, painted black on a white ground and in graded sizes are employed. The subject is seated at a distance of 6 meters (20 feet) and a figure is placed with the gap

of the **C** or the prongs of the **E** turned to the right or left, he is asked to say in which position the figure is directed. The width of the lines composing the figures and the gap in the **C**, or the spaces between the prongs

of the **E**, subtend angles of various degrees, depending on the size of the figure, when placed at a distance of 6 meters. The width of the whole figure is five times the thickness of its parts. By finding the smallest figure whose position can be recognized the visual acuity of the subject (in terms of the visual angle) is ascertained. In testing the visual acuity for the fitting of glasses Snellen's test type is most commonly employed. This test is devised upon the basis that two points or lines separated by a space having a visual angle of 1 minute can be resolved by the average normal eye. The test type comprises nine rows of block letters printed in black upon a white card. The rows are arranged in descending order of size from above down. The width of the lines forming the letters of the first row subtends an angle of 1 minute at 60 meters from the eye, while that of the letters in the other rows, two to nine, have a visual angle of 1 minute at 36, 24, 18, 12, 9, 6, 5 and 4 meters, respectively. The card is placed in a good

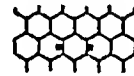


FIG 73 14 See text.

light, the patient is seated facing it at a distance of 6 meters and asked to read down as many rows as he can. The visual acuity is expressed as a fraction, the numerator being the distance at which the subject is seated from the card and the denominator the distance at which the letters could be read by the normal eye. Thus, if he reads the seventh row of letters, i.e., those with a visual angle of 1 minute at 6 meters his vision is $\frac{6}{18}$ or normal. If, on the other hand, he can see distinctly only as far as the fourth row, which the normal eye can read at 18 meters, his vision is $\frac{6}{18}$, if as far as the third row his vision is $\frac{6}{24}$, and so on, for any other row which he is just able to read.

Knowing the distance of the nodal point of the eye from the retina and the visual angle, the size of the retinal image of an object can be calculated (p. 1147). With a visual angle of 1 minute the space on the retina separating two point images is 4.4μ . The diameter of a foveal cone is given by different observers as between 2.5 and 4.0μ . Even if the higher of these figures is taken, then the image of two dots separated by 4.4μ would fall upon two cones separated by a single unstimulated cone or by one stimulated differently, i.e., by the image of the interspace (fig. 73 14).

From such calculations it has been argued that cone diameter is the limiting factor in discriminating

⁹ The minimum separable is analogous to two point discrimination in cutaneous sensation (p. 936).

two points or thin lines, for obviously with an interspace less than a cone width the two dots would fall upon a single cell, and from what is known of the nerve impulse it cannot be admitted that two parts of a visual receptor upon receiving simultaneously different types of stimulus can give rise to dissimilar sensations.

Difficulties stand in the way of so simple an explanation. The eyes, even with the most exact fixation, are constantly executing fine movements, a fact which precludes the possibility that the retinal image stimulates any set pattern of cones, the image must be constantly shifting its position. This theory must also make the assumption that when the angular distance separating two points is about the diameter of a cone the two images must be dodged about with almost incredible precision, so that they come to lie not on adjacent cones (see fig 73 14) but on two cones separated by an unstimulated cone or by one stimulated differently. Adler concluded from his experiments, in which fixation as exact as possible was secured, that a point image moves over from 2 to 4 cones at least. Furthermore, the size of the image on the retina cannot, owing to the diffusion of light, be calculated with the precision implied by the foregoing calculations (see below). It

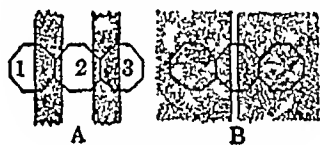


FIG 73 15 Description in text.

is true that the smallest visual angle recorded ($44''$) for space between two visually discrete objects would give a retinal image larger (3.2μ in diameter) than some of the estimations (3.0μ) of cone diameter—a fact which fits the theory that the latter is the limiting factor. Nevertheless, it is probable that the correlation is no more than a coincidence.

Wilcox and Purdy suggest that the essential factor is the total illumination of the central cone as compared with its fellow on either side. Thus, in fig 73 15 A, cone 2 receives more light than cones 1 and 3. There is no reason then theoretically why the interspace separating two lines could not be narrower than the width of a single cone and still be recognizable. A limit would be reached, however, when the white interspace was reduced to about half the width of a cone, then all three cones must be illuminated equally. It may be asked, what essential difference is there between the minimum visible (p 1117) and the minimum separable, that the minimum visual angle should be so much greater in the one instance than in the other? Two parallel black lines upon a white surface are recognized as separate because a third white line is seen between them. Why then, provided it is bright enough can it not be seen, even though reduced to an almost infinitesimal width? The difference in the two instances

appears to be a matter of the background and may be illustrated by fig 73 15 B. A bright light upon an extensive background illuminates a single row of cones, while all cones for a distance on either side are unilluminated. According therefore to the conception of Wilcox and Purdy, no difference exists, in so far as the fundamental retinal process is concerned, between the minimum visible and the minimum separable.

As with other faculties of the eye, no absolute and constant value can be given for the minimum separable. It varies greatly with several factors, viz, (a) the intensity of illumination of the test object, (b) the spectral character of the light, (c) the region of the retina stimulated and (d) the size of the pupil.

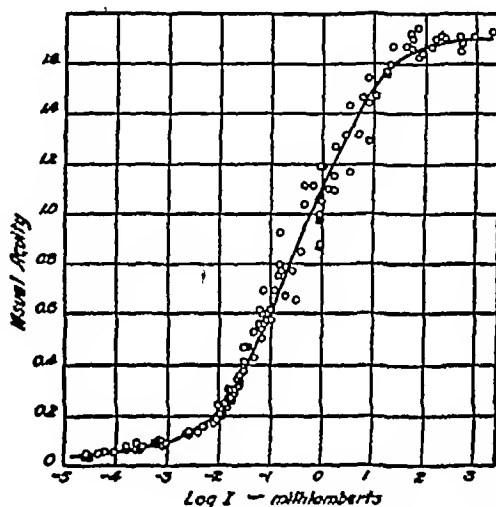


FIG 73 16. Relation between visual acuity and illumination (After Hecht)

The resolving power of the eye for two points increases with the illumination, as illustrated in fig 73 16. This is a fact difficult to explain. It might be thought that as long as the illumination was above the threshold for a just perceptible sensation, increase in the intensity of the light would be without influence upon the threshold for the minimum separable. Two theories which have been proposed to account for the phenomenon will be mentioned, not because they give any final answer to the problem, but because they bring up some interesting points.

Harttridge proposes a theory based upon the aberrations of the optical system of the eye and the ability of the retina to discriminate between small differences of light intensity (p 1121). The retinal image is not clearly defined but is blurred by diffusion circles or bands, due to the diffraction of light at the pupillary

margin and to colored fringes (chromatic aberration, p 1160) Assuming the pupil to be 3 mm in diameter and a foveal cone 3.2 mm across, he made the following calculations of the light distribution on the retina caused by a white line separating two dark areas. The illumination of the row of cones corresponding to the center of the line image was taken as 100. The illumination of the next row was 31%, of the next 9%. The diffusion bands thus virtually increase the width of the image of the white line beyond that indicated by calculations from the visual angle, and encroach upon the dark boundaries. The fine movements of the eyes are continually shifting the image and even the slightest movement will cause the line of junction between the outermost diffusion band and the dark area to move from one row of cones to another. A row of cones is

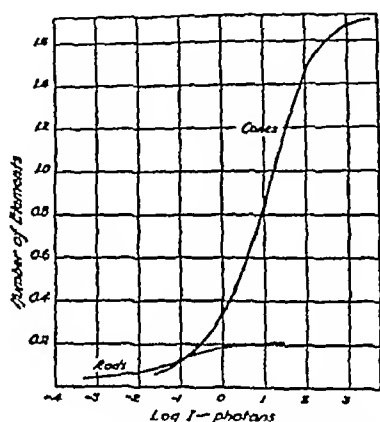


FIG 73.17 Statistical distribution of sensitivity of rods and cones. The ordinates read directly in units of visual acuity (After Hecht.)

therefore stimulated at one instant at a different intensity than at the next. If the difference in light intensity is greater than 10% it is appreciated, and thus a gap between the two dark areas is detected. Now the threshold for discrimination of differences in light intensity varies with the illumination (p 1122). Consequently, when the illumination is reduced the difference between the intensity of the outermost diffusion bands and of the dark areas is not perceived, i.e., the shift of the line of junction from one cone to the next causes no sensation. The difference between the illumination of the inner and outer bands may be still detectable, but the width of the bright line will be reduced. With further reduction in the illumination and the consequent raising of the threshold for intensity discrimination the white image becomes still narrower and finally, when the angular width is about 1", the dark areas fuse across it.

Hecht's theory postulates a change in the "grain" of the retina as a result of variations in illumination. The retina, according to Hecht, is made up of sensitive elements of different intensity thresholds distributed in

a statistical manner similar to that of other populations. At low levels of illumination only a proportion of these elements (i.e., rods with the lowest threshold) are excited. The active elements are therefore farther apart and the "grain" of the sensitive surface is relatively coarse. As the light intensity increases the thresholds of more and more elements are exceeded. More rods function, that is, the number of active elements is increased and the "grain" of the retina is finer. As the illumination rises further, cones, first those with the lowest thresholds and later less sensitive ones, become active. The maximal effect of increasing the illumination upon the resolving power is reached when the entire rod and cone population is responding (see fig 73.17). Objections to both of the theories just outlined could be cited, several observations indicate that the threshold for the minimum separable is not dependent entirely upon retinal factors but that central processes play an important part.

The resolving power of the eye is greater for monochromatic light than for a mixed light source such as daylight when the two have equal illuminating values. This fact is due to the absence of chromatic aberration in the former instance. Monochromatic yellow light (575 mμ) gives the highest value, next in order come green, red and blue. Mercury arc light, owing to its greater homogeneity, gives higher values than ordinary white light.

Three factors are involved in the effect of pupillary size upon the resolving power of the retina. Increase in diameter allows more light to enter the eye and therefore increases the illumination of the retinal image and raises the visual acuity. Diffraction is also reduced by widening the pupillary aperture which will tend to improve the definition of the image. On the other hand, narrowing the pupil diminishes chromatic aberration. The optimum pupillary size lies between maximum constriction and full dilatation, namely, at a diameter of about 3 mm.

With ordinary illuminations the visual acuity is some twenty times greater at the fovea than in any outlying part of the retina. In the dark adapted eye (p 1124) the peripheral retina has a much higher value than the fovea.

The illumination of the field surrounding the test object (the surrounds) has an important influence upon the resolving power of the eye. A uniform increase in the illumination of the surrounds up to $\frac{1}{10}$ that of the test object progressively increases visual acuity. Raising the illumination of the surrounds from this point to equality with the test object causes a slight reduction

in visual acuity and, when the surrounds become brighter than the test object, there is a decided depression. A very bright but small light source situated in the neighborhood of the test object, e.g., a motor head light, causes a very marked lowering of the visual acuity. The effects caused by such concentrated sources of light are referred to as "glare." If the small light source is not too bright and especially if the surrounds are dark, little depressing effect upon the acuteness of vision is produced, indeed there may be an improvement due to the accompanying pupillary constriction.

THE DISPLACEMENT THRESHOLD OR THE VERNIER ACUITY These terms are applied to the visual faculty of recognizing a break in the contour of a border, a variation in width of a line or the lack of alignment of two straight lines placed end to end. This power of the eye is some ten times greater than its ability to resolve two points. A break in a line subtending an angle as small as $5''$ or even $1''$ can be detected under optimal conditions. It seems quite certain that this visual faculty is not limited by cone diameter, for the break must lie on a single cone, and the lines on both sides of the break on the same row of cones (fig 73 18). It is probable that the underlying mechanism is different from that governing the threshold for the discrimination of two points. For example, its threshold is only slightly raised by increasing the illumination (p 1122).

Anderson and Weymouth offer an interesting theory to account for the extraordinary accuracy of the vernier acuity. They suggest that the slight but continuous eye movements shift the line image over the retina, causing successive stimulus patterns. The averaging of the successive patterns gives a sense of position which they call *retinal local sign*. The longer the lines the greater are the number of patterns presented to consciousness and, consequently, the more accurate is the averaging process. For details of the view of these authors the reader is referred to their original paper.

SOME PRACTICAL CONSIDERATIONS WITH REGARD TO LIGHTING Besides reducing visual acuity, glare causes discomfort and one instinctively attempts to protect the eyes by closing the lids or raising the hand as a shield, the pupil constricts. A constant source of glare, even of mild degree, results in eye strain. Glare has been classified into three types—veiling, dazzling and blinding. *Veiling glare* is that due to strong light which, being uniformly superimposed upon the retinal image, reduces contrast. The light reflected from a printed page under a bright sky is an example. *Dazzling glare* is due to scattered light in the ocular media which

does not form part of the retinal image. Such glare can be produced by a strong light shining into the eye from an angle of about 45° . *Blinding glare* results when one looks directly at a very bright light. It is due to an actual reduction of retinal sensitivity.

For moderately fine work, such as reading, sewing, typesetting, etc., the illumination of the objects should not be less than from 10 to 20 foot candles. An illumination of 10 foot candles is sufficient for reading ordinary black type on good paper, but the higher illumination is necessary if the printing or the paper is of poor quality. The effect of lighting upon the performance of typesetters, mail sorters and others engaged in fine work has been the subject of a number of investigations. Raising the illumination has been found to increase the rapidity and accuracy of the work by from 10 to 16 per cent and to reduce eye strain and general fatigue. The maximum efficiency ap-



FIG 73 18 (After Adler) See text

pears to be reached when the illumination is about 20 foot candles. The lighting should be diffuse, and naked bright light sources which could cause glare eliminated. The central field should receive additional lighting so that its illumination will be from 5 to 10 times that of the surrounds. The constant use of the eyes in poor lighting leads to ocular strain and fatigue with consequent headache. It may cause increase in pulse rate and even nausea, ultimately serious eye defects, especially in the young, may result. The quality of the light is also an important factor. The nitrogen light, being more homogeneous, gives a higher visual acuity than the ordinary electric light bulb, the kerosene lamp lies between the two.

THE DISCRIMINATION OF DIFFERENCES IN LIGHT INTENSITY This faculty was first investigated by Bouguer in 1760 and later by Weber (1834) and Fechner (1858). The latter observer found that when the light intensity is gradually increased the least change in illumination which can be perceived by the subject occurs in a series of steps. This relationship is expressed in the Weber-Fechner law which is applicable not only to vision but to other senses as well. The law states that the least per-

ceptible difference between a series of stimuli is in each instance a certain constant fraction of the preceding stimulus. For example, let us suppose that for vision the fraction is 100 and the initial illumination is 100 candles. Then, if the light of one more candle is added, the difference in illumination will be recognized, that is, the light of 101 candles is perceived to be greater than that of 100. If there were 1000 candles to start with, 10 more would need to be added before any difference could be noticed. If, therefore, the logarithms of the intensities are plotted (abscissae) against the least perceptible differences in sensation (ordinates) a straight line should result. The law may be stated in another way, namely, that in order to cause a series of equal increments in sensation the

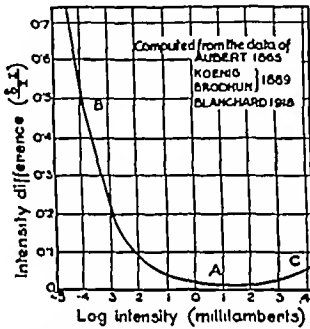


FIG 73.19 Curve of the intensity discrimination of light. (After Hecht.)

strength of the stimulus must increase in geometrical proportion. Or again, the added intensity (ΔI) necessary to cause a just perceptible difference in sensation (ΔS) bears a constant ratio ($\Delta I/I$) to the preceding intensity (I). Thus, $\Delta S = K (\Delta I/I)$

When the logarithms of the light intensity are plotted against the just perceptible differences, the curve shown in fig. 73.19 is obtained. The curve is composed of three parts. It is only section A representing the relationship at moderate light intensities that accords even approximately with the Weber-Fechner law. At intensities below and above this limited range marked deviations from the law occur (sections B and C). The whole curve covers a range from the lowest intensity to one of dazzling strength and comprises 572 steps of just perceptible differences. Starting with the very lowest intensities, e.g., 0.0000484 millilamberts (I) the least absolute increment of the stimulus which is effective (ΔI) is very small (Koenig and Brodhun's data), namely, 0.000031 millilamberts, and increases progressively as the intensity rises, so that at an illumination of 147.0 ΔI has a value of 3.803 ml. The fraction

$\Delta I/I$, therefore, is not constant, nor does it show a continuous change in one direction, for it diminishes from about $\frac{1}{3}$ at the lowest intensity to about $\frac{1}{10}$ at moderate ranges and then rises again to $\frac{1}{10}$ at the highest intensity.¹⁰ Hecht considers that the low intensity part of the curve (B) represents rod function and the high intensity portion (C), the activity of the cones. The cones are therefore more sensitive to differences in illumination than the rods, though their threshold for light perception (intensity threshold) is higher.

INTERMITTENT RETINAL STIMULATION, FLICKER
When the retina is stimulated intermittently by a series of light flashes as may be produced by interrupting a continuous light by a rotating notched disc or by reflecting light from a rotating disc divided into alternate black and white sectors, a characteristic flickering or unpleasant glittering sensation is experienced when the periodic stimulation reaches a certain frequency. This is due to each light stimulus falling upon the retina during the time of the positive after image (p. 1140) of its predecessor. The suppression of the after image by the second stimulus causes the first sensation to end more abruptly, and, through contrast, to render the succeeding one more brilliant. Upon further increasing the speed of rotation and, in consequence, the frequency (number per second) of the light stimuli, fusion results and the flicker disappears to be replaced by a continuous sensation having a brilliance equal to the mean of the two (bright and dark) impressions, the frequency at which this occurs is called the *critical fusion frequency* (CFF). If at the instant of fusion the illumination of the bright patch in millilamberts be designated a , its area designated b , and the total area of the disc, c , then the sensation produced is equal to that which would result from a continuous stimulus having the value $(a \times b)/c$. This is known as the Talbot-Plateau law. It accounts for the well-known fact that a gray sensation of any depth can be matched by throwing black and white images alternately and at a suitable frequency upon the retina. Similarly white and red images give, at the critical fusion frequency, a sensation of pink, blue and red a sensation of purple, yellow and green of yellowish green, and so on. These effects are simply explained by *visual persistence*, that is, the sensation evoked by one stimulus has not ceased before the next one

¹⁰ Some recent observations indicate that this increase with higher intensities does not occur if the eye is fully light adapted, the fraction remaining at the lowest value even though the illumination is unpleasantly intense (see Craik).

is produced, thus a blend of the two sensations in consciousness results. The law holds true only for moderate light intensities.

The value of the CFF is variable, depending upon several conditions the most influential of which is the intensity of the light, the value rising as the intensity increases. That is, a higher rate of stimulation is required for fusion as the intensity of the illumination is increased. The influence of light intensity is embodied in the Ferry (1892)-Porter (1906) law which states that the *critical fusion frequency is directly proportional to the logarithm of the light intensity*. Thus, $n = k \log I + k'$ where n equals flashes per second at the instant of fusion, and I the light intensity, k and k' are constants involving the size of the stimulated area and the sensitivity of the observer's eye.¹¹

The Ferry-Porter law is valid, however, only under certain special conditions, it holds over moderate ranges of illumination of the test object when the image is restricted to the fovea. Above and below this middle range the linear relationship between the logarithm of the intensity and the critical fusion frequency does not hold. When the value of n is plotted against $\log I$ at low and at high intensities the points fall on two straight lines, one at low the other at high intensities. It is believed that these represent respectively rod and cone function, a conception borne out by the results of Hecht and associates and of Lythgoe and Tansley.

Hecht found that with a stimulus restricted to the fovea (cones) the relationship was linear for a middle range of illumination, but above this the curve flattened out, below, it formed a very gentle curve. With the image on extra-foveal regions the data form two intersecting straight lines, one at lower the other at higher intensities, the former presumably represents rod function, the latter peripheral cone function (fig. 73 20).

Lythgoe and Tansley observed that during dark adaptation the CFF falls in both the fovea and the peripheral parts of the retina when the intensity of the test light was high (6.8 foot candles). At low intensities (0.020 foot candles) the CFF also falls at the fovea, but rises in the peripheral retina. Now, as judged by other criteria, only cones are functioning at high intensities whether the fovea or the peripheral retina is being tested, at low intensities only rods. Also, it was found by Lythgoe and Tansley that when red light was used for testing (rods insensitive), and in a case of night blindness (defective rods) a fall in the value of the CFF occurs during dark adaptation. The fall in the

value with high illuminations of the test object is due presumably to the cones, and the rise with low illuminations, to the rods. At moderate illuminations of the test object a fall occurs during the first 5 minutes of dark adaptation (due to cones), (see p. 1113) followed by a rise (due to rods). The critical frequency due to the rods is highest with dark surrounds, that due to cones is increased by bright surrounds, the maximum being reached when the brightness of the latter and of the test object are equal.

A study of the retinal potentials (ch. 74) during intermittent stimulation shows that when a light flash falls upon the retina during the "off effect" of a preceding stimulus, the d wave is interrupted and a pronounced negative dip occurs. This is an exaggerated a deflection (Pm). The negative deflection is followed

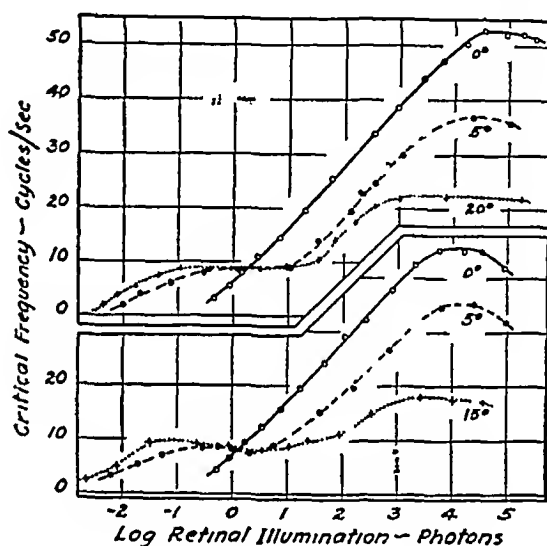


FIG. 73 20 Relation between critical frequency and $\log I$ for white light with a 2° field in four retinal locations, at the fovea, and at 5° , 15° , and 20° above the fovea. (After Hecht from Hecht and Verrijp.)

by a large positive swing which, if the second stimulus occurs soon after the first, is simply the return of the momentarily interrupted "off effect" of the first. If the interval between the two is greater, the upswing is higher, and is then due to the b deflection (Pm) of the after coming stimulus. Thus, if the light flashes are so timed that each interrupts the "off effect" of its predecessor, a series of regular ripples appears in the electroretinogram which apparently are the cause of the flickering sensation (fig. 73.21). No negative dip occurs nor do the characteristic ripples appear if a flash falls upon the retina *before* the "off effect" of the preceding one—a continuous sensation should therefore result.

ADAPTATION OF THE EYE TO DIFFERENT LIGHT INTENSITIES (See also p. 1113)

The retina behaves differently in partial darkness than in bright light. It has been pointed out else-

¹¹ Determination of the CFF offers an accurate and convenient method for comparing the brightness of differently colored lights. It is especially valuable in this regard because of our natural tendency to confuse the brightness of a color with its hue or saturation.

where that this is attributed to a difference between the rods (of one type) and the cones

DARK ADAPTATION SCOTOPIC VISION The phenomenon was first studied by Aubert in 1865. It is common experience that upon passing from light into darkness we become blind for the moment, but soon a faint glimmer of light appears and gradually more and more of the detail of our surround-

It has been doubted that dark adaptation of the fovea occurs, this is now conceded though it is of minor degree. When the eye is fully dark adapted the sensitivity of the fovea is only about 1000 that of the extrafoveal retina. It is a familiar fact that in partial darkness an object may be seen "out of the corner of the eye" (peripheral vision) but is invisible when the eyes are turned to look

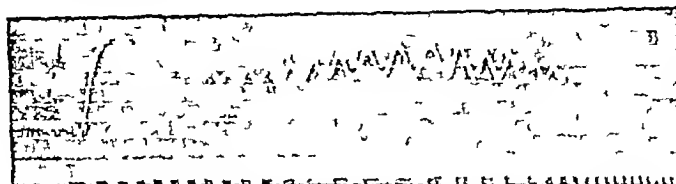


Fig. 73.21 Electroretinogram recorded during intermittent stimulation. See text. (From Granit.)

ings becomes dimly visible. The adaptation of the eye to the low illumination is associated with the following changes:

- (1) Dilatation of the pupil
- (2) Increased sensitivity of the retina, i.e., lowering of the intensity threshold and of the thresholds for the minimum separable and the discrimination of intensity differences
- (3) Purkinje shift (p. 1113) towards the blue end of the spectrum
- (4) Regeneration of visual purple in the rods
- (5) Certain structural alterations in the retina (of cold blooded animals) and a change in reaction from acid to alkaline (see p. 1130)

In bright illumination the pupillary diameter is about 3 mm, but when the eye is shielded from light the pupil immediately commences to dilate and reaches a diameter of 8 or 9 mm in full dark adaptation. In the dark adapted eye the first detectable response—slight constriction—of the pupil occurs at an illumination of about 0.025 meter candles.

The intensity threshold of the peripheral retina commences to fall almost upon the instant that the eye is darkened, the fall being very rapid for the first 10 minutes. After this and up to 30 minutes, the fall in the threshold is less precipitous and from then on is very slow. For practical purposes the extrafoveal retina may be considered to have reached its maximum sensitivity after about 40 minutes in the dark, though a gradual slight increase occurs for 90 minutes thereafter. The light sensitivity of the peripheral retina is increased by from 10,000 to 20,000 times by dark adaptation (fig. 73.22).

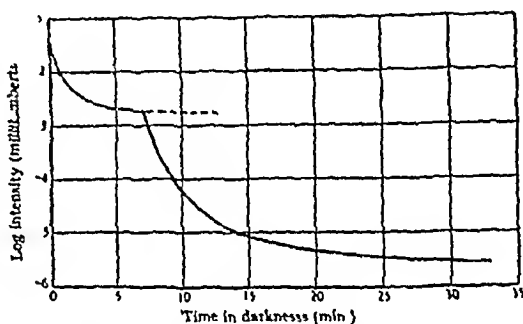


Fig. 73.22 Dark adaptation curve. The dotted line—continuation of the upper section of the curve—is the cone adaptation curve obtained when a centrally fixated red stimulus is used. The lower part of the curve indicates rod adaptation. Only the upper part of the curve and its extension is obtained in night blindness. (After Rawdon Smith.)

directly at it (foveal vision). Foveal dark adaptation is very rapid, being almost complete within from 3 to 4 minutes. It is represented by the first part of the curve in figure 73.22. In night blindness only this part of the curve exists. Hecht found that a small red cross (wave length greater than 650 mμ), to which the rods are almost insensitive, was seen at a much lower threshold of illumination by the dark adapted eye. The threshold was found to fall from 450×10^{-4} ml to 10×10^{-4} m lamb in 30 seconds and to 3×10^{-4} in 20 minutes. This type of curve is also obtained when a small, white test object is imaged on the center of the fovea.

The color sense in the dark adapted eye The rods are insensitive to red rays and give no color sensation to most other wave lengths. In the dark adapted eye they have a much lower threshold for white than has the fovea; therefore, a colored ob-

ject with an illumination value around the threshold of the rods first appears without color. As the illumination of the object is increased its color is perceived when the threshold of the cones is reached. The interval between the two thresholds is called the *colorless (achromatic) interval*. The colorless interval for deep red (λ greater than 620 m μ) is very small, since the rods (like the photographic film) are only slightly sensitive to it or not at all. In other words a red object appears black in a very dim light and is detected with difficulty against a dark background, not until the intensity of its illumination reaches the threshold of the cones is it perceived. In red light, therefore, dark adaptation of the rods is not prevented. This fact was applied during the last World War. Airmen for half an hour or so before a night flight remained in a room illuminated by a deep red light, bright enough to enable them to read, play cards, etc., yet which would permit dark adaptation to proceed as in darkness. A more convenient method is the use in an ordinarily lighted room of goggles fitted with dark red glass. Adaptation for night vision was found to be as complete when induced in this way as when carried out in darkness. The achromatic interval is longest for blue. In severe instances of night blindness the Purkinje shift and the achromatic interval for all wave lengths are absent, this is what one would expect to result from the rod defect.

LIGHT ADAPTATION PHOTOPIC VISION The dark adapted eye is dazzled by even moderately bright light, but adaptation to the higher illumination develops rapidly. During light adaptation changes opposite in nature to those occurring in darkness take place, namely,

- (1) Constriction of the pupil
- (2) Diminished sensitivity of the retina
- (3) Purkinje shift towards the red end of the spectrum
- (4) Bleaching of visual purple
- (5) Shift in reaction of the retina from alkaline to acid and structural changes, in certain species (p 1130)

The eye adapts much more readily to light than to darkness. In its measurement the threshold at full adaptation is first determined, the eye is then exposed to bright light for a certain period after which the light is switched off and the threshold found before any appreciable degree of dark adaptation has developed. This procedure is repeated for different time intervals following the preliminary period of dark adaptation. It is found that the greatest decrease in sensitivity

occurs during the first 20 or 30 seconds, this is followed by a gradual fall for a period of 10 minutes. After this time light adaptation may be taken for practical purposes to be complete, though some slight reduction in sensitivity occurs up to 30 minutes.

THE PERCEPTION OF MOVEMENT This is the most primitive of the visual functions, in disease it is the last to fail and is the first to return should any improvement in vision occur. The peripheral (extrafoveal) retina is a specially differentiated organ for the perception of movement (Exner) being much more sensitive in this respect than is the fovea. It is a familiar fact that a slight movement is more readily detected if the moving object is not in the direct line of vision, i.e., when its image falls upon the peripheral retina, than when the eyes are fixed upon it. The most sensitive part of the retina is from 10° to 15° from the fovea, but diminishes progressively towards the periphery. In the region of maximum sensitivity the angular velocity of a just perceptible movement is from one half to one minute per second, provided that there are stationary objects in the visual field to serve as reference points. When such are absent the angular velocity must be from 10 to 20 times as great in order for the movement to be perceived. On the other hand, if the angular velocity is very great the movement is not perceived, owing to visual persistence a very rapidly moving object appears as a stationary streak. The total distance travelled, i.e., the displacement of the object, as well as the angular velocity is, of course, a factor in movement perception. The minimum displacement is about 17 seconds of arc, under optimal conditions. The sensitivity of the retina to movement is lower in the dark adapted than in the light adapted eye.

When the eyes are stationary but the body or head is moved an *apparent* movement is given to objects in the visual field. When travelling in a train, for example, near objects appear to move in the opposite direction to the direction of travel, while those in the background appear to move with the moving vehicle. Apparent movements of surrounding objects also occur when the eye is displaced slightly by pressure upon it with the finger tip, or as a result of involuntary contraction of the eye muscles. These apparent movements are attributed to the successive stimulation of groups of receptors as the images move over the retina. When the eyes are moved voluntarily from one object to another in the visual field, images must sweep over the retina in a similar fashion, yet there

is no apparent movement of stationary objects.¹² Conversely, the movement of an object is perceived when it is followed by the eyes, though the position of its retinal image does not alter. It is quite evident that the perception of movement is very complex and cannot be explained in all its aspects upon physiological grounds. It is suggested that the absence of an apparent movement of

images over the retina. In other words, the successive stimulation of visual receptors is ignored because the point to which the eyes are to be turned engages the attention at the moment that the eye movement takes place, or even before. The perception of the movement of an object pursued by the eyes must also depend upon cerebral processes.



FIG. 73.23 Illustrating irradiation.

stationary objects when the eyes are turned from one part of the visual field to another is to be explained upon the basis of *attention*. The attention exercised by the observer in changing the fixation of his eyes from the one to the other point compensates, it is believed, for the movements of the

Apparent movement is also produced by the stimulation of closely approximated retinal areas in rapid succession by a series of images of a stationary object. The two main factors determining this so-called *stroboscopic illusion* of movement are the time interval between the stimuli and the *angular separation* of the successive retinal images. A visual sensation of smooth motion is produced when the angular separation is about 1 degree or less and the intervals between the stimuli about $\frac{1}{16}$ second. At intervals of $\frac{1}{8}$ second or less no sensation of movement is produced. The illusion of motion is also lost if the time intervals are lengthened to $\frac{1}{4}$ second or greater, the impressions then becoming discrete.

Irradiation. Owing to chromatic and spherical aberration the images on the retina are not formed of geometrical points of light, but rather of bright points surrounded by diffusion circles. For this reason, and also probably as a result of the spread of the effect of the stimulus (*irradiation*) to neighboring neural elements of the retina, or even within the visual area in the brain, a bright area on a dark ground appears larger than a dark one of the same size upon a bright ground. In either instance the image of the bright area encroaches upon that of the black area (see fig. 73.23).

¹² An allied phenomenon and one which offers a similar problem to be solved is seen in cutaneous sensation. We are able, for example, to distinguish between the movement of the finger over a stationary object and the movement of an object over a motionless finger. In both instances receptors are stimulated successively.

THE NATURE OF RETINAL PROCESSES, OBJECTIVE RETINAL PHENOMENA, COLOR VISION, OPTICAL ILLUSIONS

THE NATURE OF THE RETINAL PROCESSES

The photochemical mechanism of vision. Visual purple or rhodopsin is a rose pink photosensitive pigment found in the outer segments of the rods (upon which it is probably adsorbed), it has not been demonstrated in the fovea. The pigment was discovered by Boll in 1876 and extracted from the retina by Kühne in 1878. Its presence is essential for normal vision in dim light, and to it the nervous layer of the fresh retina, detached in the dark from the pigment layer, owes its pink color (p. 1110). It can be extracted from the retina by bile or a solution of bile salts, but the most effective extracting agent is digitonin. Visual purple either in situ or in solution is bleached by white light or by light of any wave length between 385 m μ and 600 m μ but very slightly at these extremes (see figure 74 2, page 1129). Since it is pink in color with a tinge of blue, visual purple, as is to be expected, transmits more of the longer and the shorter wave-lengths (red and blue) than of those in the mid part of the spectrum (yellow and green). These latter are, therefore, absorbed to the greatest extent, a photosensitive substance is acted upon only by the rays which it absorbs. Maximal bleaching is caused by green light with a wave length of about 500 m μ . In the living animal and also in solution, the pigment is regenerated in the dark or in dim light. For its regeneration in solution a substance derived from the epithelial cells (pigment cells forming the outermost retinal layer) but which is not the brown pigment itself, is essential for the process. This substance, it seems, is consumed in the reaction, for when a solution of visual purple is repeatedly bleached and regenerated, less and less rhodopsin is produced.¹ For the regeneration of visual purple in solution, the pH is of prime importance. The reaction proceeds most readily at a pH of 6.7, below 6 or above 7.5 little pigment is formed.

Though visual purple is bleached by strong light and reformed in the dark, a moderate amount of light exerts a favorable effect upon the regenera-

tion process, for after bleaching, more of the pigment is produced with moderate illumination than in complete darkness. The short-wave-lengths are most effective in this respect, and the experiments of Chase and of Chase and Smith suggest that a blue-sensitive (i.e., a blue-absorbing) substance is responsible for the favorable action. It may be a yellow intermediate product formed during the bleaching reaction. The regeneration of visual purple is not affected by section of the optic nerve, but the drug santonin retards it, and in large doses causes yellow vision and night blindness. When one eye is illuminated, the other being shaded, bleaching of rhodopsin occurs only in the former.

If the eye of a rabbit is excised in darkness or in red light (light of a wave length greater than 650 m μ does not bleach the visual purple) and then exposed to an object clearly defined in light and shade, e.g., a window sash against the sky, an image of the object will be found to have been impressed upon the retina when it is examined in the dark room. The image is caused by the bleaching of the visual purple where the bright parts of the image fell upon the rods. The retina thus behaves like a photographic film, the image so obtained is called an *optogram* (see fig. 74 1).

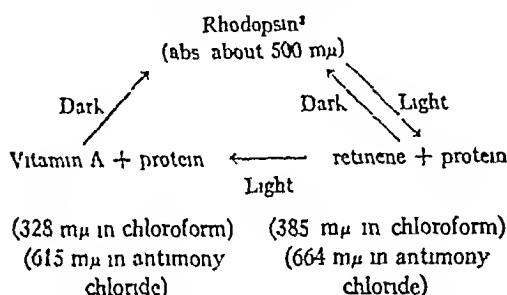
Our knowledge of the chemistry of rhodopsin has been greatly advanced within recent years. It is a complex conjugated protein containing a carotenoid pigment as a prosthetic group. The latter is called the *chromophore group*, and is the specific part of the molecule. About ten chromophore groups are believed to be attached to each protein "base" or "carrier." One quantum of light energy causes a chemical change in one chromophore group. The protein is also linked with phospholipids. The molecular weight of visual purple has been variously estimated at values ranging from 100,000 to 800,000. The most generally accepted figure is that of 270,000 obtained by Hecht and Pickels.

When visual purple is bleached by light, either in the retina or in solution, it first turns an orange color which changes rapidly to yellow and finally becomes colorless. According to Wald, rhodopsin breaks down during bleaching into protein and

¹ A substance contained in the epithelial cells was described many years ago by Kühne and claimed by him to be essential for the regeneration of visual purple. He called it *rhodophyllin*.

an orange-yellow carotenoid pigment which he calls *retinene*. Retinene is extractable by chloroform in which it shows a broad absorption band maximal at $387\text{ m}\mu$. In antimony trichloride it shows maximal absorption at $664\text{ m}\mu$. The final colorless product is vitamin A. An extract of completely bleached retinas, therefore contains no rhodopsin but is rich in the vitamin. Under physiological conditions, vitamin A does not accumulate in the retina but is carried away in the circulation. Regeneration of rhodopsin from vitamin A takes place in the retina of the living animal during dark adaptation.

The rhodopsin-vitamin A cycle is shown in the following scheme, modified from Wald



* Lythgoe has given the name *transient orange* to the first product of the bleaching process and *indicator yellow* to the second derivative. The former is unstable and decomposes spontaneously into the latter. Indicator yellow, as its name implies, behaves like an acid base indicator, being chrome yellow between pH 3.3 and 4.0, a deep lemon yellow at neutrality and a pale lemon between pH 9.3 and 10.0. Indicator yellow is also formed immediately from rhodopsin by heat and by the action of acids and alkalis. The mixture of yellow pigments formed during bleaching was called "visual yellow" by Kühne. The relationship between Wald's retinene and Lythgoe's indicator yellow is not altogether clear. They appear to represent different stages in the bleaching process. In indicator yellow, it is suggested by Lythgoe, the chromophore group is still attached to the protein "base".

² The photosensitive pigment in the retinas of fresh water fish differs from that found in the retinas of most vertebrates including the marine fishes. This pigment, which is truly purple in color, has been named *porphyropsin* by Wald. It follows a retinal cycle identical with that of rhodopsin except that vitamin A₂ takes the place of vitamin A (A₁), and the retinene produced during bleaching has, in chloroform, an absorption band with a maximum at $405\text{ m}\mu$ instead of at $387\text{ m}\mu$. Wald has made the highly interesting observation that the type of pigment present in the retinas of the euryhaline fishes (those species which migrate from salt to fresh water or vice versa) is determined largely by the waters salt or fresh, in which they are spawned. Thus, in fish such as salmon, which spawn in fresh water (anadromous forms) the predominant pigment is porphyropsin, whereas, in those, such as, the fresh water eel which spawn in the sea (catadromous forms) rhodopsin predominates.

Retinene has been shown more recently to be vitamin A aldehyde.

In the completely bleached and isolated retina, or in the intact eyes of animals fed upon a diet deficient in vitamin A, regeneration of rhodopsin cannot take place, though in the former some regeneration from retinene may occur in darkness. It will be recalled that vitamin A deficiency is one cause of night blindness (nyctalopia).

Convincing evidence can be cited for the view that the first phase in the retinal process whereby radiant energy is converted to nerve impulses (electrical energy) is photochemical in nature. There can be little doubt that rod vision is dependent upon such a process and that the photochemical pigment, rhodopsin plays an essential rôle in twilight (scotopic) vision. When the absorption



FIG 741 Optogram formed upon the retina of a rabbit by exposing the eye to an object made of a glass plate and strips of black paper (after Stewart)

spectrum of this pigment is plotted (reciprocals of the minimum energy required for bleaching along the ordinates, and the wave lengths along the abscissae) the graph is found to be identical in shape with the visibility curve for scotopic vision (fig 742). A curve formed by plotting the rate of bleaching of visual purple against the wave length coincides almost exactly with the absorption curve, in other words, visual purple follows Draper's law which states that chemical change is produced in a photosensitive material only by those waves which are absorbed. The significance of these facts is unmistakable.

Though the visibility curve of the dark adapted eye and the absorption curve of visual purple are identical in shape, there is a difference of $7\text{ m}\mu$ in the maxima of the two curves, the maximum of the former lying at $510\text{ m}\mu$ (Hecht and Williams), that of the latter at about $503\text{ m}\mu$. An explanation for this discrepancy will be offered presently. The retina also obeys the Bunsen-Roscoe law formulated for photochemical processes in general,

namely, that for the production of a given photochemical effect a constant quantity of energy is required which can be distributed within certain limits by varying either the illumination or its duration. In other words, the product of the illumination (I) and the time of exposure (T) is constant, $\text{Energy} = kIT$.

The difference between the maxima of the absorption curve of visual purple and the scotopic visibility curve is actually somewhat less than that given above. Dartnall and Goodeve point out that the visibility curves should be plotted using an equal quantum spectrum (the quantum of light energy varies with the wave length) and not an equal energy spectrum (p 1114). When this is done the scotopic luminosity curve and the absorption curve of visual purple are almost identical. Hecht and Williams have pointed out that a small discrepancy may well be accounted for by the fact that the absorption curves of rhodopsin have been determined of necessity upon *solutions* of the pigment and some difference should be expected between curves so obtained and the absorption of visual purple as it exists in the retina.

Important evidence for the photochemical nature of the initial stages of the visual process has been obtained by Hecht through the study of a much simpler photochemical mechanism, namely, that of certain invertebrate forms. The body surfaces of a clam, *Mya*, of an ascidian, *Ciona*, and of a mollusc, *Pholus*, possess light sensitive areas. When exposed to light the syphons of *Mya* after a long latent period become retracted. Several points of similarity between this reaction, which is undoubtedly photochemical in nature, and vision have been demonstrated. *Mya* shows light and dark adaptation, its sensitivity being increased in the dark and reduced by a stay in daylight. The reaction follows the Bunsen-Roscoe law, and the Weber-Fechner law is obeyed over a limited range of moderate light intensities.

The photochemical theory, though satisfactory for rod vision, failed to explain daylight vision, for no photosensitive pigment had been demonstrated in the cones of the mammalian retina. Nevertheless, as a result of his studies, Hecht considers it highly probable that foveal vision is also dependent upon a photochemical mechanism. The visibility curve of the light adapted eye is identical in *shape* with that for twilight vision and consequently with the spectral absorption curve of a solution of visual purple. But the *maxima* of the three curves do not coincide, that of the light

adapted eye being at $550 \text{ m}\mu$. The identical shapes of the curves, nevertheless, suggest very strongly that cone vision is also dependent upon a photochemical substance—possibly visual purple in high dilution. A low concentration of the pigment in or around the cones would account for its not being evident, yet its function would not necessarily be abolished thereby, for it has been shown that certain photochemical substances, e.g., silver nitrate and cyanin, do not lose their properties when spread in very thin films. However, the difference between the maximum of the daylight visibility curve and that of the absorption curve (about $45 \text{ m}\mu$) were considered too great to be accounted for purely upon the basis of changes in the properties of the pigment during its extraction.

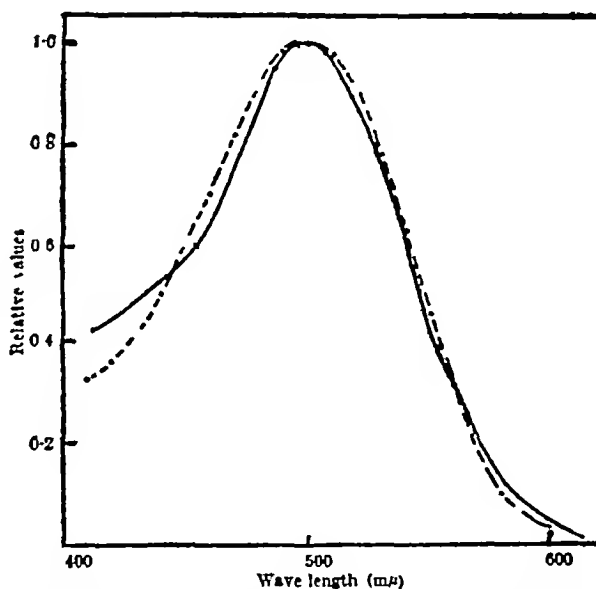


FIG 74.2 Comparison of the absorption curve of pure visual purple (interrupted line) with the scotopic luminosity curve (continuous line) as obtained with an equal quantum spectrum (After Ludvig)

The results of Wald's researches have largely resolved the difficulty of extending the photochemical theory to cone vision. He extracted the dark-adapted retinas of chickens which contain chiefly cones. Such an extract would contain rhodopsin derived from the small rod population mixed with cone pigment if such existed. Upon exposing the extract to red light, to which the rods are only faintly sensitive, bleaching occurred, absorption being maximal at $575 \text{ m}\mu$. Subsequent exposure of the extract to white light caused bleaching characteristic of rhodopsin with maximal absorption around $508 \text{ m}\mu$. The initial bleaching by red light was due, presumably, to a

photosensitive pigment derived from the cones. The spectral properties of this pigment indicate that it is violet in color. Wald has therefor named it *iodopsin* (visual violet). A comparable pigment has been obtained by Chase from the retinas of frogs. The maximal sensitivity of chicken cones is at 580 $m\mu$ and that of the rods at about 520 $m\mu$. The former figure is in close agreement with the maximal absorption (575 $m\mu$) of *iodopsin*.

NERVOUS MECHANISMS The photochemical mechanism constitutes the means whereby light energy, through the breakdown of rhodopsin, causes a visual stimulus. But this is merely a preliminary phase in the retinal process which is essentially nervous in nature. It must always be borne in mind that the retina is an outgrowth of the brain. Functionally it is a "true nervous center" (Cajal). Unlike any other sense organ its receptors are in close association with synaptic connections, and many of the phenomena of the retina are expressions of interaction between neurons. For example, phenomena which have already been described for nervous centers, e.g., spatial and temporal induction, inhibition, rebound, etc., have been demonstrated for the retina. Adrian and Matthews demonstrated spatial summation, using the shortening of the latency of the optic nerve discharge as the criterion of increase in the retinal response. They found that as compared with the illumination of a single area simultaneous stimulation of four separate retinal areas caused a marked shortening of the latent period. Granit and Davis have shown by means of the electroretinogram the occurrence of temporal summation in the human retina. They found that when the retina is stimulated by two subliminal stimuli separated by an interval of about 50 m. sec. summation to the threshold results. The effect of the first stimulus persists, though gradually declining, for 135 m. sec., after this interval, a second stimulus produces no response unless it is of threshold value itself. The retina, then, must not be looked upon simply as a mosaic of separate points, like auditory or touch receptors, each connected to the nervous centers by its own nerve fiber and acting independently of the rest (Adrian and Matthews).

OBJECTIVE RETINAL EFFECTS CAUSED BY LIGHT STIMULI

Illumination of the eye causes (a) bleaching of the visual purple, (b) alteration in reaction of the retina from alkaline to acid, (c) histological changes in the pigment cells and cones of certain

cold-blooded species, and (d) changes in electrical potential. The bleaching of visual purple has been dealt with in the preceding section.

Little is known of the chemistry underlying the production of acid in the stimulated retina, but the acid is probably lactic derived from the breakdown of glucose, for it has been shown that the retina as compared with other tissues has a strong glycolytic action. The acidity increases with the intensity of the stimulus and is greatest with yellow-green light.

Histological changes in the retinas of frogs, fish and certain other cold-blooded forms occur upon stimulation, but have not been demonstrated in mammals. They consist of movements of the pigment granules of the cells of the outer retinal layer (p. 1110) and retraction of the bases of the cones from the pigment cells, leaving an interval between the two. The contraction of the cones is apparently a reflex phenomenon, for it occurs upon stimulation of the opposite eye or even, according to some, upon exposure of the skin to light. Swelling of the rods in bright light, and then shrinkage in the dark so that they become separated by a distance of from 0.5 μ to 0.8 μ are other retinal phenomena seen in these species.

RETINAL ACTION CURRENTS, THE ELECTRORETINOGRAM (ERG) When the cornea and the optic nerve or the posterior pole of the darkened eyeball are connected through a galvanometer, a current is set up with the cornea as the positive pole. A steady deflection of the galvanometer results. The cause of this resting potential is not clearly understood. When a light is thrown into the eye a series of potential changes is produced which can be recorded as a corresponding sequence of waves. These are the retinal action currents, the record is called an *electroretinogram*. A steady current is also set up when the inner and outer surfaces of the eyeball are connected (fig. 74.3).

Holmgren, the Swedish physiologist, was the first (in 1866) to demonstrate retinal action currents when the eye is stimulated by a beam of light. From then to the beginning of the present century they were studied by a number of investigators, including Devar and McKendrick, Waller and Gotch. They were recorded with the string galvanometer by Eanthoven and Jolk in 1903 and by Piper, by means of the capillary electrometer, in a series of studies from 1903 to 1911. In recent years with the development of more delicate methods of recording the subject has been reinvestigated by Chaffee, Bowie and Hampson, by Hartline and by Granit and his associates.

In the earlier experiments records were obtained from the surviving excised eye, and in many instances after removal of the anterior half of the globe, or from

preparations in which dissection had caused a considerable degree of trauma. Though it was necessary to place one electrode directly upon the retina in order to prove the retinal origin of the currents,⁴ better results can be obtained by leaving the eye in situ and placing one electrode upon the cornea of the illuminated eye and the other (indifferent electrode) upon the eye of the opposite side or upon any moist surface of the body. Typical electroretinograms are obtained in this way. Hartline has applied this method to the human subject by means of water-tight goggles fitted to the eyes and filled with a 0.9% saline. The goggle in front of the eye to be illuminated holds the differentiated electrode and is fitted with a convex lens to correct for the optical effect of the solution. The indifferent electrode is fitted into the other goggle. In another method employed by Hartline for man, a cotton-tipped electrode is applied directly to the illuminated eye after anesthetization with a 2% solution of holocaine, the indifferent electrode is placed in the mouth.

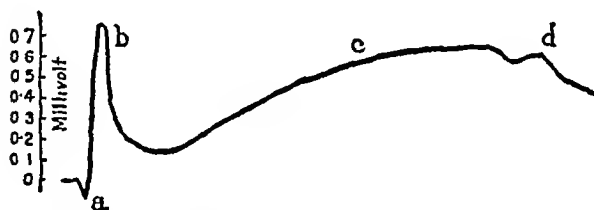


FIG 74.3 Electroretinogram (From Granit, redrawn)

The most satisfactory preparation from which to record the electroretinogram is the decerebrate animal, since perfect immobilization is thus secured, as well as full pupillary dilatation. The indifferent electrode is placed upon brain tissue on the proximal side of the decerebration cut, and the other electrode on the cornea. The recording instrument most commonly employed is the string galvanometer or some type of oscillograph, the currents being first conducted through a valve amplifier.

Analysis of the record The electroretinogram (see fig 74.3) consists of four waves designated *a*, *b*, *c* and *d*. The *a* wave is a small, rapid negative deflection—a sharp notch below the base line—which appears after a short but variable latent period. The wave *b* is a well-marked positive deflection of about 0.7 millivolts, and *c* is a slow, prolonged positive wave which declines gradually though the eye remains constantly illuminated. The *d* deflection is a small positive change which appears after a brief latency when the light is cut off, for this reason it is usually referred to as the “off effect”.

⁴ A characteristic record is not obtained after removal of the nervous layers of the retina.

Einthoven and Jolly concluded from an analysis of their records that the “on effect” (*a*, *b*, and *c* waves) was a composite deflection due to the algebraic summing of the potential changes of three separate retinal processes. The recent researches of Granit and his associates have confirmed and extended the work of these earlier investigators. The three processes have been designated P_I , P_{II} and P_{III} , since they can be made to disappear in this order by gradually increasing the depth of anesthesia. P_I is responsible for the sustained potential of the *c* deflection, it is abolished by light anesthesia, leaving the fast initial waves (*a* and *b*) unchanged. With deeper anesthesia the *b* wave becomes smaller and then disappears, leaving the P_{III} component responsible for the *a* deflection. This is a prolonged negative deflection.

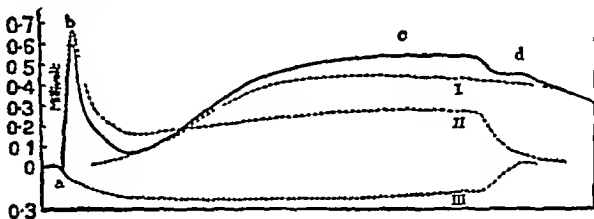


FIG 74.4 Analysis of composite retinal electrical response (cat's eye) at two intensities, 14 ml and 0.14 ml and area of 1661 sq mm, viewed at a distance of 70 mm. Components, broken lines. Composite curve drawn in full. The *a*-wave is broadened slightly out of scale to show its derivation more clearly (Granit, *J. Physiol.*, Vol LXXVII, 1933)

In the final stage of anesthesia the negative component disappears and the retina, having apparently suffered an irreversible change, cannot be made to respond again. The P_I component is only seen at high light intensities and in the dark adapted eye, P_{III} is absent, as a rule, at low levels of illumination. P_{II} is abolished by asphyxia. Thus, by varying the illumination, by anesthesia or asphyxia, records can be obtained composed of $P_{II} + P_{III}$, of $P_I + P_{III}$, of P_{II} alone or of P_{III} alone.

The three processes responsible for the electrical changes in the retina arise in the nervous layers and the potential changes do not represent nerve fiber responses since they cannot be obtained by placing the differentiated electrode upon the optic nerve entrance, *synaptic connections are essential for their production*. The *b* wave, for example, increases in amplitude with the illumination intensity, whereas, as will be seen presently, the magnitude of the action potentials of the optic nerve does not vary with the strength of the stimulus. Furthermore, the latency of the retinal response diminishes as the intensity of the light or the size of the illuminated area on the retina increases—additional evidence of synaptic transmission.

P_{III} (component responsible for *a* wave) is, according to Granit, an inhibitory process. P_{II} (responsible for *b* wave) is excitatory in nature and is followed by a

discharge of impulses in the optic nerve, and is probably a rod and cone effect. P_1 (component causing elevation) is not associated with a discharge of impulses in the optic nerve, its origin is obscure. The "off effect" (d wave) which occurs when the light stimulus is withdrawn is considered to be a rebound phenomenon (p 951) resulting from the cessation of the inhibitory process P_{III} and the release with consequent enhancement of the slower processes P_1 and P_{II} , particularly of the latter.

ACTION POTENTIALS IN THE OPTIC NERVE The action currents in the optic nerve of the conger eel were investigated by Adrian and Matthews. The optic nerve of this species is unusually long and contains relatively few fibers—facts of considerable technical advantage. The potentials were recorded by a capillary electrometer after passage through a four valve amplifier.

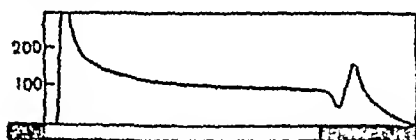


FIG 74.5 Curve of frequencies of impulses in optic nerve (After Adrian and Matthews)

The nerve is almost "quiet" in the dark but a light flashed into the eye causes a discharge of impulses. If the nerve is crushed between the eye and the recording electrodes no potential waves appear. The total response to a flash lasts for about 15 milliseconds. Depending upon the intensity and area of illumination, the latent period varies from 0.1 to 0.5 sec., becoming shorter with increasing strength of stimulus. The frequency of the impulses show two maxima, one (300 per sec) associated with the "on effect" of the electroretinogram, the other with the "off effect" (see fig 74.5).

The amplitude of the impulses (action potentials) or their grouping does not vary with the intensity of the stimulus or with the area of the retina illuminated (p 922). On the other hand, an increase in either of these factors increases the impulse frequency and shortens the latency of the nerve response. Hartline has recorded the responses of a single optic nerve fiber of the horseshoe crab (*Limulus*) and of the frog. He also found that a change in frequency of the impulses, but not in their amplitude, resulted from varying the intensity of the stimulus. Responses were evoked over a great range of light intensities,

namely, from 1 to 100,000. Enlargement of the area of the retina which was illuminated but maintaining a constant intensity of light, or conversely, with a fixed area and increasing the intensity, was followed by a rise in the impulse frequency. Thus, it was again demonstrated that the determining factor is the total amount of luminous flux (area \times intensity, p 1117).

With short flashes it was found that the impulse frequency varied with the product of the duration of the flash and its intensity, when the duration and intensity were varied reciprocally the impulse frequency remained practically unchanged (see fig 74.6). The effect of varying the

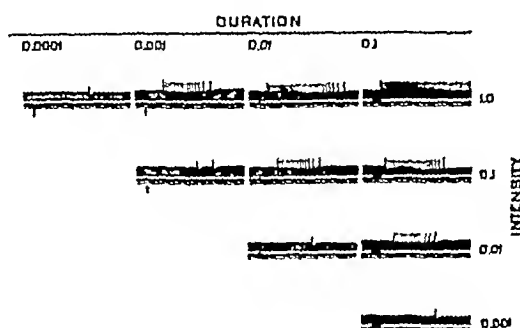


FIG 74.6 Oscillograph records of impulses from a single optic nerve fiber of *Limulus* caused by brief flashes of light of varying intensity and duration. Note that the frequency of the impulses is a function of the product of these two variables, when the products are equal the frequency of the impulses is virtually the same (after Hartline).

area is limited to images under 1 mm in diameter, the greatest effect being observed with variations in the size of images of 0.3 mm in diameter or less. Though, as just stated, the latent period (*i.e.*, the time elapsing from the application of the stimulus to the appearance of the first impulse in the nerve) is variable, the retinal nerve time (*i.e.*, the interval between the deflection of the electroretinogram and the commencement of the impulse discharge in the nerve) is constant with varying strength of stimulus. Variations in latency of the nerve response are therefore due simply to corresponding variations in the latent period of the retina.

In the peripheral retina the receptive field of a single nerve fiber as determined by exploring the retina with a small spot of light was found to have an area of about 1 sq mm. This indicates the convergence of many receptors upon a single ganglion cell (p 1109), for an area of this size contains several hundred rods and cones. The

maximal response was obtained from the center of this area

There appear to be three types of fiber in the retina of the frog, for three types of response were obtained (a) A series of single uniform spikes during constant illumination which rise in frequency as the intensity of the light is increased (b) A burst of impulses at the beginning of illumination and another burst when the eye is darkened, but no response during illumination (c) A series of impulses discharged for a second or two when the light is cut off but no response to illumination

A somewhat different method has been employed by Granit to record impulses from single optic fibers. A micro-electrode is inserted with the aid of a micromanipulator into the retina of an eye which has been deprived of its cornea and lens. When single uniform spikes are observed in the record similar to those obtained by Hartline, the impulse discharge is taken to be in a single nerve fiber. This method is suitable for the investigation of the mammalian eye.

As mentioned above, Granit's analysis of the retinal response shows that process II alone is responsible for the initiation of a discharge in the nerve. The impulses therefore reach their maxima where this process causes the *b* deflection and at the "off effect" (*d*) where rebound occurs as a result of the cessation of the inhibitory process (*P_{III}*).

The burst of impulses at the "off effect" (*d* wave) is of special interest, such is not seen in the nerve fiber of any other receptor organ. Its significance, however, seems clear. The signalling of darkness to the centers in the brain is of just as great importance in the life of the animal as the signalling of light. A shadow or a dark object in the field of vision may indicate food or the approach of an enemy.

The decline in frequency of the impulses, i.e., between the two maxima (fig 74.5), though the stimulus persists is difficult to explain. It might be attributed to sensory adaptation (ch 63). Yet how can such an explanation be reconciled with our own sensations, for, as a rule, an object at constant intensity of illumination does not appear less bright while we continue to look at it. On the other hand, it is certainly true that a moving image, which stimulates groups of retinal elements successively, is seen more easily and is more likely to attract attention than a stationary one. Indeed in many species an immobile object apparently is not perceived. Is it that in the *intact* and *conscious* animal, a sustained sensation of brightness of a stationary object under constant illumination is due to the incessant involuntary eye movements which, by causing successive stimulation of groups of retinal elements, nullify the effect of adaptation and that the optic nerve transmits a continuous stream of impulses at high frequency?

COLOR VISION

Color perception is a function of the light adapted eye (p 1125), and has been generally regarded as a function of the cones. This view is based mainly upon the facts that (a) at low light intensities when rods alone are functioning, the eye is incapable of color discrimination, as the illumination is gradually reduced red objects first lose their color, appearing black (absence of any stimulating effect), then those which in daylight are colored yellow, green or blue appear gray in the order given, (b) in bright light the maximum perception of color is within the foveal area, where rods are few or absent. Therefore, animals whose retinas contain only rods are thought to be completely color blind.⁵

But there are certain observations which argue against this relatively simple and clear cut conception. The central fovea, as first pointed out by König and recently emphasized by Willmer is blue-blind. A small blue object, for example, when viewed from a distance of a few feet, so that its image falls well within the fovea appears gray or black. Only when the object is brought close to the eye and its larger image then overlaps the outer foveal area and the surrounding peripheral retina does it appear blue.⁶ This and other evidence suggests that the rods are concerned with the perception of blue. In blue light, for example, visual acuity is lower at the fovea than in the peripheral retina. The blue cast given to white objects in deep twilight or in moonlight is thought to be dependent upon the rods. The sky at night, so long as there is any light at all, is not gray or black, but a deep blue. It has been postulated that a blue-sensitive substance (p 1110), possibly a yellow intermediate product formed in the bleaching of visual purple, is contained in the rods (see Duplicity Theory, ch 73).

Color possesses three qualities or attributes—hue, brightness and saturation.

Hue or tone is dependent entirely upon the wave length of the rays. Thus, red, yellow, green, blue and violet are different hues.

Brightness (brilliance or luminosity) is determined by

⁵ Many insects, some bony fishes, reptiles, most birds and the primates are capable of discriminating color. The dog, cat, guinea-pig, rat and most other laboratory animals are color blind.

⁶ Hartnidge, however, states that this phenomenon is not confined to the central fovea but can be demonstrated in the peripheral retina.

the intensity of the rays,⁷ whatever their wave length, falling upon the retina. Thus, there can be many degrees of brightness within a single hue. Of two objects having the same hue and illuminated to the same degree, the one which absorbs the smaller proportions of the rays of the specific wave length and reflects the greater proportion appears the brighter. The different brightnesses of a given hue may be compared to a number of mixtures of black and white made up in various proportions to produce a graded series of shades of gray. For example, just as any shade of gray may be produced by adding black to white, so the brightness of a paint of any given hue can be reduced to the degree desired by mixing with it a suitable quantity of black. The brightest colors are those near the middle of the spectrum—the yellow and green. It is very difficult for one to match the brightness of two different colors (heterochromatic photometry). The method mentioned in footnote p 1122 is resorted to, which overcomes the natural tendency to confuse brightness with hue and saturation.

Saturation (or purity) As in the case of brightness, color saturation may show infinite variations in degree within a given hue. Red is a more saturated color than pink. The difference is due to the relative quantities of white light with which the red light is mixed. For example, the saturation of a paint of a given hue can be reduced by mixing white with it. The mixture reflects white light as well as light of the wave length by which its hue is perceived. The saturation of a colored light could be readily determined by means of the spectroscope. Thus, a fully saturated red or green (homogeneous or monochromatic) light would give a red or a green color at the corresponding part of the spectrum, whereas, if either color contained an admixture of white light, all the spectral colors would appear, their intensity depending upon the quantity of white in the mixture. The brightest colors of the spectrum are also the least saturated. The saturation of a given spectral color is expressed by the following log $((L_s - L_w)/L_s)$. Where L is the amount of spectral radiation which must be added to white of luminosity L_w to give a just noticeable change of color from white.

THEORIES OF COLOR VISION

None of the many theories of color vision is capable of accounting for all the observed phenomena, either of normal vision or of color blindness. The theory suggested by Young (1807) and supported with extensions and elaborations by Helmholtz (1853) is the best known, the simplest and perhaps open to the fewest objections.

⁷ It might be more strictly correct to say "intensity of sensation" rather than "intensity of the rays" for we have seen that dark adaptation causes a shift in the point of maximal luminosity, but as measured physically of course no shift occurs.

THE YOUNG-HELMHOLTZ THEORY Though scarcely necessary perhaps it may be an advantage to recall some elementary principles. White light is a combination of light rays of wave lengths from about 720 mμ to 350 mμ, and can be split into its constituent colors—red, orange, green, blue and violet—by means of a transparent prism. These spectral colors can be recombined again to produce white light of the same intensity as the original light from which they were derived (Abney's law). It is not necessary, however, for one to use all these colors in order to produce white, only three, namely, red, green and blue (or violet) combined in suitable proportions, are required to produce white or any other color we may desire. Red, green and blue (or violet) are therefore called the primary colors. But pure spectral lights, not pigments, (see footnote, p 1139) must be employed in the combinations. If we should set up three lanterns, of which one emitted red light, another green and a third blue, then by throwing one upon a screen or a blend of two or of all three, we could produce any color of light we chose as well as white. Red and green would give rise to yellow, red and blue or violet to purple, and so on, all three colors, mixed in equal proportions, would produce white. These are uncontested facts to which any theory of color vision must conform.

The Young-Helmholtz theory postulates three types of percipient elements (receptors) in the retina, each containing a photochemical substance which is acted upon by the rays of one or other of the primary colors. The sensation of white and of the many different colors is due to the stimulation of the three elements at different relative intensities (see fig 747). Pure red light, for example, stimulates strongly those cones containing the "red substance" and the other types very feebly, green, or blue light stimulates powerfully those containing the green or blue substances, respectively, the sensation of white results when all three types are stimulated equally.

Though only three types of cone are postulated by the Young-Helmholtz theory, the really important part of this theory is not that any particular number of cones is specified, but rather that the analysis of color is placed in the retina. For, just as the analysis of the pitch of sound is a function of the basilar membrane, so the visual receptors with different spectral sensitivities give the cue to the brain for the interpretation of color. It is possible that the cells of the lateral geniculate

body represent the segregation of impulses from the three types of retinal receptor (p 1170) Thus, one may conceive of a topical relationship between the three types of cone and the cells of the primary visual center which may be continued to the visual area of the cerebral cortex That is, "red" receptors would be connected with "red" brain cells, "green" with "green" and "blue" with "blue" It should be stated, however, that histological evidence for the existence of three types of cone is lacking But the failure to secure such evidence by no means puts the theory out of court, since differences in the chemical structure of the photochemical substance contained in the cones would escape microscopical detection

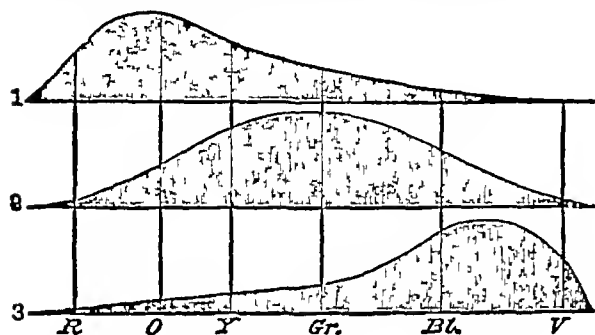


FIG 747 Diagram of three primary color sensations 1 is the so-called 'red' 2 'green,' and 3 'violet' primary color sensation R, O, Y, etc, represent the red, orange, yellow, etc, colors of the spectrum The diagram illustrates, by the height of the curve in each case, how the several primary color sensations are respectively excited to different extents by vibrations of different wave-lengths After Helmholtz)

This theory is in accord with most of the established facts in respect to color mixing with normal persons, i e, color matching with mixtures of lights of appropriate wave-lengths

It is evident that the theory is in accord with the established facts in respect to color mixing It also gives a plausible explanation of the common types of color blindness on the assumption that one or other type of cone is defective or lacking For example, failure to appreciate red would be due to deficiency of the red-sensitive substance, absence of sensitivity to green to lack of the green substance Temporary partial color blindness can be produced by intense stimulation of the retina by one or other of the primary colors, which is readily explained upon the basis of fatigue of the corresponding type of sensitive element. For example, if one looks at the sun for a couple of minutes through a red glass and then directs the eyes to differently colored objects, the perception of red is found to be lost—red geraniums appear black, yellow flowers green and pink roses sky blue (rose color being a blend

of pink and blue) The eye fatigued in a similar way by green light sees the foliage colored bluish or reddish gray

The trichromatic theory has proved inadequate, nevertheless, to explain certain observations, and a number of alternative theories have been proposed, such as Herring's, Franklin-Ladd's, Roaf's, Hartridge's and Granit's The first three of these are mainly of historical interest and will not be described

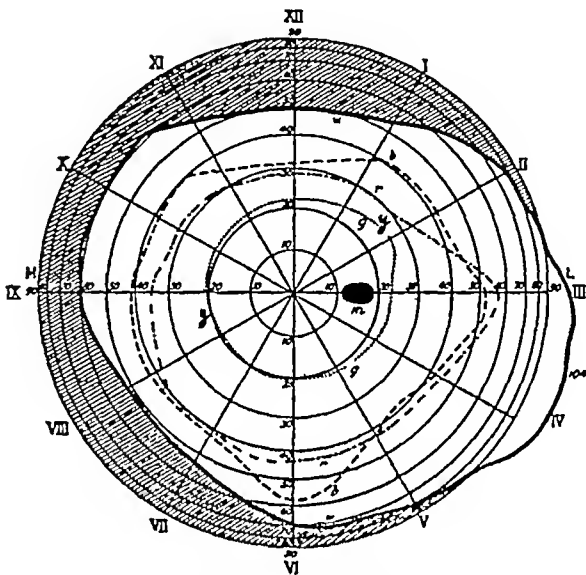


FIG 748 Showing the color fields of the human retina mapped by means of the perimeter (chap 76) (After Hartridge) When with the eyes of the subject looking straight ahead a colored patch (green, yellow, red or blue) is moved peripherally in the different meridians of the visual field the sensation of green is the first to be lost, next yellow, then red, and finally blue

Hartridge cites several objections to the acceptance of the classical three-color theory in so far as the extra foveal retina is concerned For this region he substitutes a *polychromatic theory*, postulating the existence of seven types of receptor, of which the chief are, *orange*, *green* and *violet* and smaller numbers subserving *red*, *yellow*, *blue-green* and *blue* sensations The fovea, however, he believes to be approximately trichromatic

Since the center of the fovea contains only cones, or cones and a few rods, which are believed to be of the "day" type, this part of the retina is blue blind (tritanopic)

Granit from his studies of the responses in single optic fibers concludes that luminosity and color are subserved by independent receptors which he calls, respectively, "*dominator*" and "*modulator*" units They are sensitive over a wide range of wavelengths Their response curves, constructed by plotting the energy of threshold stimulus against wave-length, closely

simulate the human photopic luminosity curve. The "modulator units" are in smaller numbers, and are sensitive to narrower bands of the spectrum, they are divisible into three main types according to the spectral regions to which they respectively respond, namely, between 580 and 600 $m\mu$ (red), 520 and 540 $m\mu$ (green) and 450 and 470 $m\mu$ (blue). Granit considers that the "modulators" by acting in some way upon the responses of the dominator units give rise to color sensations. The dominator units being sensitive over such a broad band of the spectrum can give no cue to the brain with respect to color. The fundamental or dominant visual sensation is a colorless "white" or brightness, and it is the function of the receptors with the narrowly restricted spectral sensitivities to "modulate" this dominant sensation in such a way as to give direction to the brain for the interpretation of color.

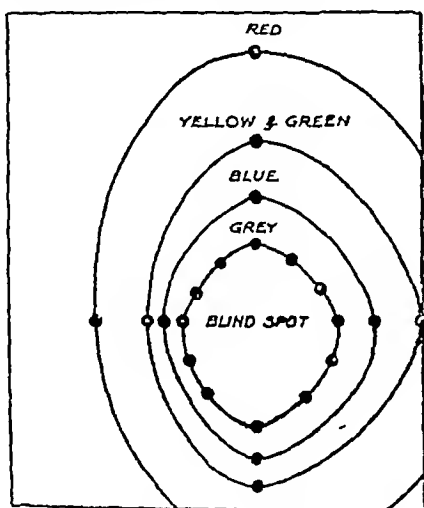


FIG 749 Color fields surrounding the blind spot. (After Haycraft)

THE DISTRIBUTION OF THE COLOR SENSE The determination of the extent of the retinal areas sensitive to the three primary colors is important in many pathological conditions, for changes in the distribution of the color sense may be one of the first changes to be detected in certain visual defects. When the normal retina is mapped out by means of the perimeter (ch 76) it is found that the three primary sensations and yellow are distributed over four areas of different sizes which are centered, roughly, at a point a little to the inner side of the fovea. The largest area is for the perception of blue, a short distance within its circumference is the boundary of a smaller area for red. Next in size is the area for yellow, the area sensitive to green is the smallest, this last

area is therefore trichromatic (sensitive to all three primary colors). A band surrounding it, i.e., the interval between the boundaries of the red and green areas is sensitive to blue, yellow and red, the third zone to yellow and blue, and the fourth to blue alone. The retina lying beyond the blue area as far forward as the ora serrata is achromatic, being sensitive only to white light, a colored object appearing gray or black. Minor variations in the extent and shape of the color fields are seen in different persons. The margins of the fields are rather irregular and include more of the nasal than of the temporal part of the retina, i.e., more of the temporal part of the visual field (p 1167). This distribution applies only to stimuli of moderate intensity. With very strong stimulation sensations of red, yellow and blue can be evoked from the periphery of the retina. In other words, the areas for white, red, yellow and blue are of about the same size. The area for green is, however, even with a light of high intensity, less extensive than that for white or for the other colors. The apparent blue blindness of the central fovea has been mentioned (p 1133) (see fig 748).

The color fields of the retina surrounding the blind spot are not of the same relative sizes as those just described for the peripheral retina. It will be seen from figure 749 that the area for red is the largest, next in order is the area for yellow and green and then the one for blue. A narrow zone beyond the latter and immediately surrounding the disc is sensitive only to white light.

COLOR BLINDNESS

We have seen that, according to the Young-Helmholtz theory, normal color vision is trichromatic, i.e., white light or any color can be matched by a mixture of three primary colors in suitable proportions. Defects of color vision are classified upon the basis of whether three, two or one of the primary colors are required by the subject to match any color of the spectrum. Thus, the main division is into *anomalous trichromatic*, *dichromatic* and *monochromatic* types. The three forms could be explained in terms of the Young-Helmholtz theory by presuming that one or other color sensitive substance—red, green or blue—was defective or absent.

ANOMALOUS TRICHROMATIC VISION In this defect, which was discovered by Lord Rayleigh in 1882, there is not complete blindness for any color, but the appreciation of red or of blue-green is less

than normal. The anomaly with respect to red is called *protanomaly* or *partial protanopia*.

Subnormal perception of green is called *deutanomaly* or *partial deutanopia*. The subject of protanomaly, in order to match a homogeneous yellow light with a mixture of red (λ 671 m μ) and green (λ 536 m μ) requires the addition of far more red than does a person with normal vision. The subject of deutanomaly requires more green. This is called the Rayleigh test. Anomalous trichromatic vision is looked upon as a transition stage between normal vision and the dichromatic form of color blindness.

The defective perception of blue—a very rare abnormality—is called *tritanomaly*. In order to match a blue-green the tritanomalous person requires more blue than the normal.

An instrument known as an anomaloscope is employed for the detection of these defects. In the case of protanomaly or deutanomaly, a monochromatic yellow light is matched with a mixture of red and green lights. The subject is asked to adjust the proportions of the two latter colors until an exact match is made. The amounts of the two colors required are shown by the instrument and compared with that required by a person with normal color vision.

DICHROMATIC VISION. There are three types of dichromatic vision, *protanopia* or *red-blindness*, *deutanopia* or *green-blindness*⁸ and *tritanopia* or *blue blindness*. For the sake of illustration we may presume that in this form one of the three photosensitive substances postulated by the Young-Helmholtz theory is lacking. Color vision is therefore a function of two variables instead of three, i.e., only two primary colors, green and blue in protanopes, red and blue in deutanopes, or red and green in tritanopes, are required to match white or any color of the spectrum.

Protanopia is the commonest form, tritanopia is very rare. To the protanope the red end of the spectrum is shortened, red objects appear dark and may be confused with dark green, dark gray, dark blue or black, the red-yellow region of the spectrum appears in different shades of green. There is no darkening of the red end of the spectrum in deutanopia. In protanopia the red-yellow part of

the spectrum appears as different shades of green. John Dalton (1798), the famous English chemist, suffered from this type of color blindness and was the first to give a scientific account of it (1794), hence the term Daltonism. He says "Crimson has a very *grave* appearance, being the reverse of every showy and splendid colour. Woolen yarn dyed crimson or dark blue is the same to me. . . the colour of a florid complexion appears to me that of a dull opaque blackish-blue upon a white ground."

It has just been stated that in dichromatic vision, white or any spectral color can be matched by the mixture of two primary colors in suitable proportions. It follows, therefore, that in dichromats, a certain mid-region of the spectrum (at a wave-length of about 493 m μ in protanopes and of 497 in deutanopes) appears colorless and is matched with light gray. This so-called *neutral point* is blue-green to the normal eye. Wave lengths within this range therefore give the impression of a dirty white and when combined with other rays have a corresponding value in the mixture.

When mixed with other colors, red (in protanopes) and green (in deutanopes) give sensations such as would result from equivalent quantities of gray. The red or green in the mixture is not perceived. As a consequence of such anomalies, pale colors within the neutral range and certain shades of red and green or compound colors containing one or other of these, are confused (*confusion colors*). The dichromat matches pale green with grays, buffs and straw colors, as well as with pale green. The red-blind also confuses yellows with greens, a bluish pink (rose color) with pale blue or violet and dark reds with dark browns, greens or blues. Red-blind persons show, as a rule, a certain degree of red-green blindness as well.

The luminosity curve (ch. 73) of the light adapted eye of protanopes differs from the normal, the spectrum is shortened toward the red end and maximal luminosity is nearer the blue. In deutanopes, the luminosity curve differs little from the normal though the point of maximal luminosity may be a little farther toward the red end of the spectrum (fig. 74.10). The luminosity curves of the dark adapted eye of dichromats and of normal persons are identical. Owing to the absence of color at the neutral point persons with dichromatic vision can detect smaller differences in luminosity than can those with normal vision. This serves the dichromat in discriminating between certain colors which otherwise would be confused, actually, as a result of the improvement in this faculty, some 140

⁸ Many authorities, especially adherents of the Hering hypothesis of color vision, refer to either protanopia or deutanopia as red-green or green-red blindness, since in both conditions red and green are confused. Similarly, tritanopia is called yellow-blue blindness.

different hues can be differentiated by such persons—the number for the normal eye is about 160

Tritanopia or blue blindness is nearly always acquired, i.e., due to some eye disease affecting the function of the rods, such as, detachment of the retina, blues and greens are confused. The neutral point is in the yellow part of the spectrum at about 572 $m\mu$. The condition is simulated in jaundice because the blue rays are absorbed by the yellow bilirubin which

tinged with blue (blue sensation of the rods). The spectrum usually shows as in normal scotopic vision, maximal luminosity at about λ 500 $m\mu$.

About 8 per cent of males and 0.4 per cent of females show some defect of color vision. Deuteranomaly is the most common type (see table 99).

Both forms of anomalous trichromatism and of dichromatism are nearly always hereditary. The condition follows a mode of inheritance similar to that of hemophilia, i.e., it is a sex-linked character carried in the X-chromosome (p. 121 and p. 871). As mentioned on page 121 hemophilia may in very rare instances occur in a girl or woman should her father be a bleeder and her mother a transmitter. This also applies to color blindness which, owing to the more frequent occurrence of the condition as compared with hemophilia, accounts for the 0.4 per cent of females affected.

TESTS FOR COLOR BLINDNESS When one remembers the important part played by red and green—the two colors which are most commonly confused—in the control of our rapid transportation systems it is unnecessary to stress the dangers of defective color vision in operators, e.g., signal men, engine drivers, and seamen. In all civilized countries, applicants for employment in the marine and railway services must pass rigid color tests. Subjects of color blindness may be quite unaware of their defect, for they have no means of comparing their actual sensations with those of normal persons. They have learned to associate their own visual sensations with certain common names, e.g., red, green, etc., but unless their peculiarity is brought home to them by some glaring error, e.g., wearing a red tie to a funeral, they may not realize that the sensation which they know as red is not the same as that experienced by others. For this reason dichromats can usually name ordinary colors correctly and depend upon their power of fine discrimination between differences in luminosity to distinguish two colors which in hue appear to them alike.

Three types of test will be described.

(1) *Spectroscopic*. In this method the extent of spectral visibility,⁹ the discrimination between the different hues, the position and extent of the neutral band and

⁹ Roaf found in the investigation of twenty-four red blind persons that the shortening of the red end of the spectrum varied considerably. In some, the red color extended to λ 620 $m\mu$, in others to λ 580 $m\mu$ and in a third group the spectrum was colorless beyond only λ 480 $m\mu$.

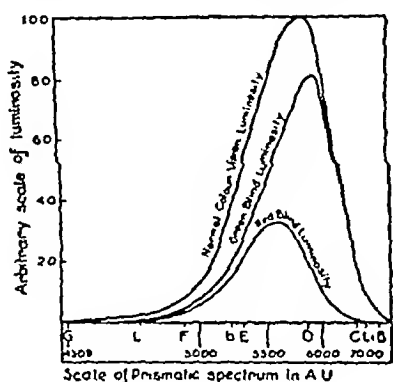


FIG. 74.10 Photopic luminosity curves of dichromats (protanopes and deuteranopes) compared with the normal trichromat. (After Abney.)

TABLE 99
(After Wright)

Type of color defect	Percentage
Anomalous trichromatism	
Protanomaly	1.0
Deuteranomaly	4.6
Tritanomaly	0.0001
	5.6
Dichromatism	
Protanopia	1.2
Deuteranopia	1.4
Tritanopia	0.0001
	2.6
Monochromatism	0.003
Total	8.2

stains the conjunctivae and ocular media. For a similar reason it is also seen as a temporary condition following the ingestion of a large dose of santonin.

Monochromatic vision or total color blindness is extremely rare. In a sense it may be looked upon as the converse of night blindness in that the function of the rods is unimpaired, but the color sensitivity of the cones is lost. To a person with monochromatic vision the surroundings appear as they do to one with normal vision at twilight, namely, in black, grays and white.

the region of maximal luminosity are investigated. The determination and comparison of the ratio of red and green required to produce a homogeneous yellow (Rayleigh test) has been mentioned (p 1137). (2) *Matching*. One of the oldest and best of the methods is Holmgren's (1877) wool test. The examinee is given a set of colored wools among which are a number of so-called confusion colors. He is asked to select from the group those which match a separate wool of a specified color. In the first stage of the test the separate wool is a pale green. If he is red or green blind he matches it with cream, buff, dove gray, pale brown and straw-colored samples. He is next given a rose pink wool to match, if he chooses violet and blue wools he is red-blind, if greens, reds and browns he is green-blind. As a final test he is handed a red stain, the protanope chooses from the mixed wools dark greens and browns, the deuteranope pale greens and browns. (3) *Pseudo-isochromatic diagrams*. Ishihara's is the most satisfactory test of this type. The test comprises a series of cards upon each of which a colored field (pale green, rose, red, etc.) is printed in spots of different sizes. A letter or figure, also made up of spots, is outlined against the general field, these spots are of a color likely to be confused with that of the field. The subject is asked to name the figure or letter, it stands out clearly to the normal eye but to the color blind may be indistinguishable from the background.



FIG 74 11 Illustrating contrast (Hering) Observe through tissue paper

CONTRAST EFFECTS

It is well known that when black is placed against white or vice versa they "set one another off", the black looks blacker and the white a purer white than if either were placed against a colored ground. Gray also appears darker against a white than against a black ground (fig 74 11). It is also true that blue against a yellow ground (or yellow against a blue ground) is more vivid than if placed against any other color. Also, green is enhanced by red and red by green. These phenomena are examples of *simultaneous contrast* or *spatial induction*.

The maximum effect of color contrast is obtained when *complementary colors* are placed side by side. Any pair of colors which when fused as lights produce white, are said to be complementary to one

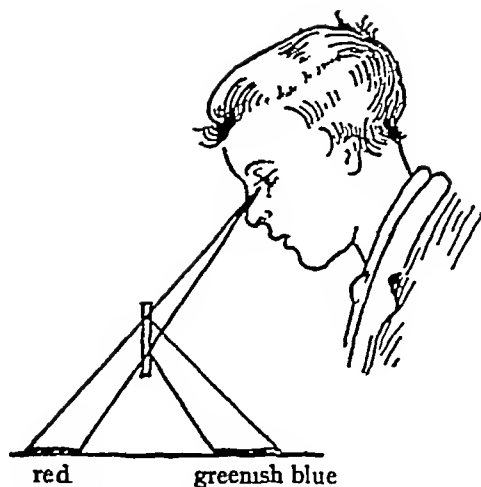


FIG 74 12 Description in footnote 12

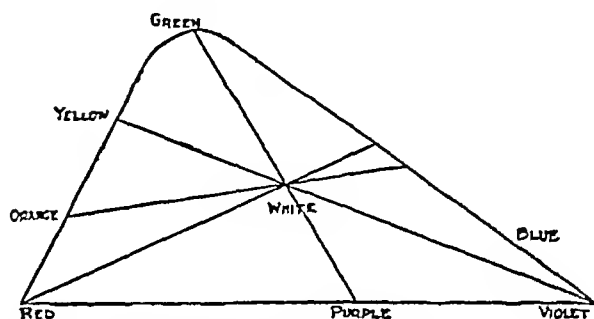


FIG 74 13 Color diagram showing complementary colors

another.¹⁰ Examples of complementary color pairs are the following

Red and greenish blue
Orange and cyan blue
Yellow and indigo blue¹¹
Violet and greenish yellow
Purple and green

¹⁰ Two sheets of paper in complementary colors (e.g., red and bluish green) or two books with their covers similarly colored are laid flat in front of the observer as shown in figure 74 12. A clear glass plate is held vertically above and midway between the colored surfaces. The far surface is seen directly through the glass and the near surface by reflection. A sensation of white is caused by the overlapping of the two images.

¹¹ A distinction must be drawn here between lights and paints, we know that yellow and blue paints when mixed give green, not white. The color of a paint is due to its reflecting certain wave lengths in white light and absorbing the remainder. Even the bluest paint is not pure blue, it absorbs all light but blue and a little green. These it reflects. A yellow paint is not pure yellow, it reflects a little green with the yellow. When the two paints are mixed, blue light is absorbed by the yellow paint, and yellow light by the blue. The green rays, of which a part is reflected by each separate paint, are doubly reflected when the two paints are mixed.

Not only those listed above, but every color and shade of color has its complementary (fig 74.13). There are consequently a great number of complementary pairs. When a color is placed in juxtaposition to gray, the gray appears tinted with the complementary of the color. For example, a lemon or other yellow object, placed in a bright light, casts a shadow which is tinged with blue—the complementary of yellow. The shadow cast by a red object becomes tinged with greenish blue, and one cast by a purple object with green. These

FIG 74.14 Optical illusion. The distance from A to B appears to be greater than that from B to C, they are the same.

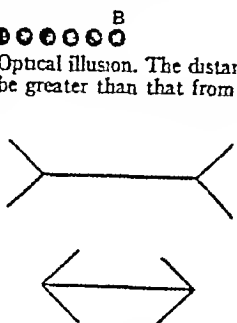


FIG 74.15 Illusion of size. The vertical lines are the same length.

principles are applied in art. The artist makes a yellow flood of sunshine more brilliant by painting blue into the shadows which in turn are given depth and an appearance of reality which otherwise they would lack.

When one stares for a time at a colored surface in a strong light and then directs the gaze to a gray surface it appears tinged with the complementary color, and objects of the complementary color itself are made more vivid. For example, if one looks at a red surface for a time and then at a green one the latter color is intensified. This phenomenon is called *successive contrast* or *temporal induction*.

AFTER IMAGES If the gaze is directed to a bright white light for a moment and the eyes then closed or turned towards a dark surface, an image of the light slowly floats into view, becomes more distinct for a time and then gradually fades. Similarly, if the eyes are stimulated by a colored light or a brightly colored object of any sort, and then darkened, an image of the same color appears. These are called *positive after images*. If, instead of closing the eyes or turning them to a dark surface after looking at a white light, the retinas are stimulated a second time and diffusely by

white, e.g., by directing the eyes to a sheet of paper, one then sees a dark image against a white ground. This is called a *negative after image*. If the first stimulus was colored, then the after image is in the complementary color. Negative after images of colored objects are the cause of successive contrast described in the last paragraph.

On the basis of Young's theory of color vision, the phenomenon of negative after images is due to fatigue of one or other of the three types of receptor by the first stimulus. Cones which have responded to a given stimulus will not for a time respond

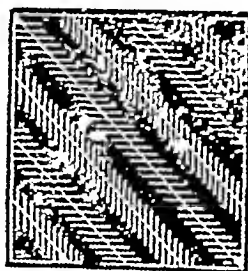


FIG 74.16 Zollner's lines. The long diagonal lines appear to converge, actually, they are parallel.

again to one of the same type. White light stimulates all three types of cone. The negative after image which appears upon applying a circumscribed and then a diffuse white stimulus to the retina is, therefore, a dark patch against a white background. When the object looked at is colored and the retina is then stimulated by directing the eyes to a white surface, the image is in the complementary color because only those cones which had not been previously stimulated can respond. For example, if the object looked at is red, the red-sensitive cones are not excited by a subsequent stimulus of white, those sensitive to green and to violet alone respond, giving a sensation the complementary of red, namely, a bluish green.

Positive after images are apparently due to chemico-physical changes in the receptors of the retina caused by and outlasting the stimulus—a form of visual persistence.

OPTICAL ILLUSION The brain may be deceived by imitations of certain effects upon which our visual judgments of the size, shape and distance of objects are based. Visual errors of this nature are called optical illusions or optical deceptions. Some interesting examples are shown in figs 74.14 to 74.16.

THE DIOPTRIC MECHANISMS OF THE EYE CATARACT. OPTICAL DEFECTS INTRA-OCULAR FLUIDS

PRINCIPLES OF REFLECTION AND REFRACTION DEFINITIONS AND TERMINOLOGY

Light falling upon a surface undergoes *absorption* and *reflection*, and, if the material is transparent, the rays are transmitted through it, either with or without *refraction*

The proportions of rays falling upon an opaque unpolished surface which undergo absorption and diffuse reflection, respectively, vary with the character of the surface. A large part of the rays striking an unpolished white surface, e.g., a sheet of paper, are reflected but, being thrown off at different angles to the perpendicular, they do not meet at a focus in front of, or, if continued backwards, behind the surface. The light reflected from such a surface is said to be *diffuse*

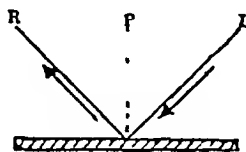


FIG 75.1

The greater proportion of the light striking a polished surface (e.g. a mirror) is reflected, but *the incident and reflected rays are always in the same plane and the angles (angles of incidence and of reflection) which they make with the perpendicular are equal*. This statement is true for any polished surface whatever its shape (fig 75.1)

Reflected rays from a plane mirror are divergent, if continued backwards they would meet at a point situated at the same distance behind the mirror as the object emitting the light lies in front of it. The eye placed in the path of the reflected rays projects them to this point, where a *full-sized erect image* is formed (fig 75.2). Since the rays do not actually meet at this point but only appear to do so, the image is called *virtual*

THE FORMATION OF IMAGES BY SPHERICAL MIRRORS
A spherical mirror is the segment of a sphere, its reflecting surface may be *concave* or *convex*, its *center of curvature* is the center of a sphere of which the reflecting surface forms a part. The middle point of the curved surface is called the *pole* of the mirror, and a line passing through the pole and the center of curvature is termed the *principal axis*. The radius of the mirror is the distance from the pole to the center of curvature. Since the latter may lie on the same side as the source of light (concave mirror) or on the opposite side (con-

vex mirror) the radius may be *positive* or *negative*, respectively

Rays of light coming from a distant object, i.e., from infinity, are *parallel* (1 and 2, fig 75.3), if they fall upon a *concave mirror* they are reflected as converging rays and meet in front of the mirror at a point (F) on the principal axis (p-o). This point (F) is the *principal focus* and the distance from it to the reflecting surface is the *focal length* or *focal distance* of the mirror. It lies in the principal axis, midway between the center of curvature and the surface of the mirror. A *real inverted image* of the object and *smaller* than it, is formed in front of the mirror at the principal focus, that is, in space. The rays from a near object are divergent, the reflected rays are therefore less strongly convergent than when the

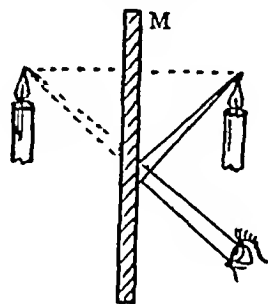


FIG 75.2

incident rays are parallel. If the object is at the center of curvature (C) of the mirror then the rays are reflected back to this point. When the object is at a between the center of curvature of the mirror and the principal focus the rays meet beyond the center of curvature at A, when the object is at A the reflected rays meet at a. These two points are therefore reciprocally related and are called *conjugate foci*. An object placed between the mirror and its principal focus emits rays which upon reflection are widely divergent and cannot be brought to a focus in front of the mirror. Projected backwards they meet at a point behind the mirror where a *virtual erect* and greatly *enlarged* image is formed

Parallel rays striking a *convex mirror* are reflected as *divergent* rays which if continued backwards would meet behind the mirror at the principal focus. To the eye they therefore appear to come from this point. Here a *virtual erect image*, smaller than the object, is formed (fig 75.4)

The position of an image formed by a concave mirror can be found from the construction in figure 75.5. The object AB is situated beyond the center of curvature C. The rays AP and BO parallel to the principal axis

MN after reflection meet and cross at the principal focus F. The images of the points A and B, therefore, lie somewhere on the lines PQ and OL. Now, if lines AD and BH be drawn to pass through C, these lines, known as secondary axes, will cut lines PQ and OL at a and b respectively. Thus, a small inverted real image ab is formed. The image is called real because the rays actually pass through a and b. The construction of an erect virtual image by a convex mirror is

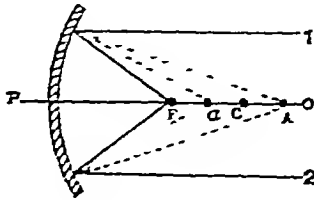


FIG 753

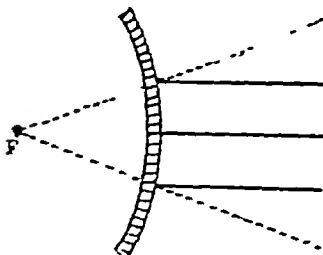


FIG 754

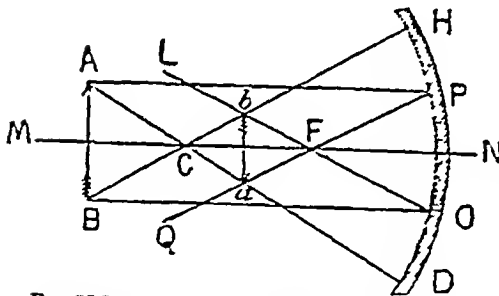


FIG 755 Formation of a real inverted image by a concave mirror

shown in figure 756. The rays from the object A B diverge after reflection and appear to come from F (dotted lines), the image ab is formed at the intersection with the secondary axes AC and BC.

REFRACTION Rays of light in passing obliquely from one transparent medium to another of a different optical density, (e.g., from air to glass) are bent or refracted. If one medium is surrounded by the other (e.g., glass in air) the ray is refracted twice. In passing from the medium of lower to the one of higher optical density the rays are bent towards the perpendicular,¹

¹That is, they become more nearly perpendicular

in the transition from the denser to the rarer medium the bend is away from the perpendicular. With any two media the greater the obliquity of the incident rays the greater is the degree of refraction, rays perpendicular to the surface between the two media are not refracted. The ratio of the angle made by the incident ray (i.e., the ray falling upon the surface of the second medium) with the perpendicular (angle of incidence) to that made by the emergent ray (angle of refraction) is termed the *index of refraction*. The *refractive index* is the index of refraction when the incident rays enter a substance from a vacuum. In practice air is considered

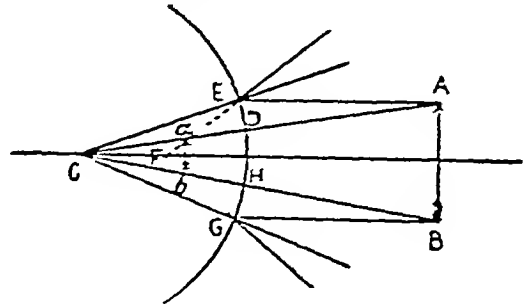


FIG 756 Formation of a virtual erect image by a convex mirror

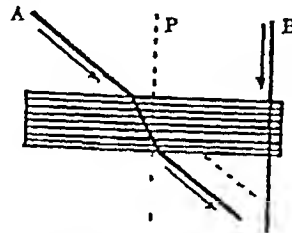


FIG 757 The oblique ray (A) is refracted upon entering and leaving the block of glass. The emergent ray has the same direction as the entering ray but is not in the same line. P represents perpendicular. The ray B which strikes the glass surface perpendicularly is not refracted.

to be of the same optical density as a vacuum. The refractive index is expressed as the ratio of the sine of the angle of the incident ray, (i) to the sine of the angle of refraction (r), thus, refractive index = $\sin i / \sin r$. The refractive index is a measure of refractive power, it is 1.52 for crown glass and 1.66 for flint glass.

Refraction by plane surface. Oblique rays striking a medium with plane parallel surfaces, such as a sheet of glass, are refracted to an equal degree upon entering and emerging, but in opposite directions, i.e., towards and away from the perpendicular, respectively. The incident and emergent rays are therefore parallel though not quite in the same straight line (fig 757). A prism has its sides inclined towards one another. Since a ray is refracted upon entering at one surface of the prism and again upon emerging at the other, and is

bent so as to run more nearly perpendicular to the glass surface in the former instance and away from the perpendicular in the latter, it will be refracted each time towards the base of the prism (fig 75 8)

Lenses are of two main types, *spherical* whose surfaces are the segments of spheres and *cylindrical*. There are six varieties of the former—*planoconvex*, *biconvex*, *planoconcave*, *biconcave*, *convexoconcave* and *concavoconvex* (see fig 75 9)

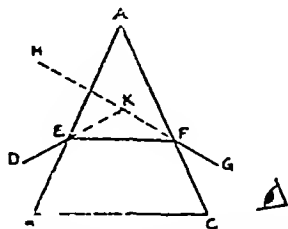


FIG 75 8 ABC is a prism with the apex at A, the base BC, and the sides AB and AC. The angle of the prism is BAC. A ray of light DEFG is refracted at E and F as in figure 66 8. The total amount of refraction, that is, the difference in direction between DE and FG, is represented by the angle DKH (the angle of deviation). If the eye is at G, the source of light, D, will appear to be at H. When the ray passing through the prism (EF) is parallel to the base (BC), the ray is said to traverse the prism symmetrically.

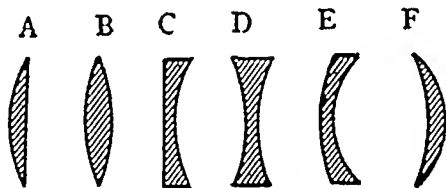


FIG 75 9 Cross sections of lenses A, planoconvex, B, biconvex, C, planoconcave, D, biconcave, E, convexoconcave, F, concavoconvex

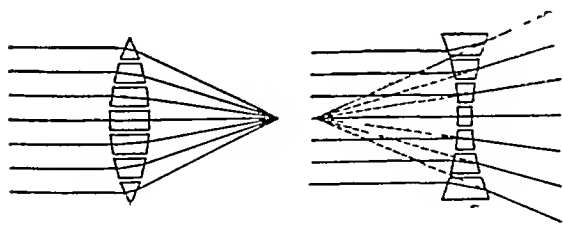


FIG 75 10 See text.

Convex lenses may be looked upon as a great number of truncated prisms with their bases directed towards the lens center. Concave lenses, on the other hand, are as a number of prisms arranged with their bases towards the periphery (fig 75 10). It follows then that a symmetrical biconvex lens will bend rays to the same degree at equal distances from its center and bring them to a meeting point or focus, whereas concave lenses will cause divergence of the rays. In either instance the rays are bent towards the bases of the constituent prism.

Refraction by convex lenses A line passing through the centers of curvature of the lens is termed the *princi-*

pal axis of the lens. Any other line intersecting the principal axis within the lens (i.e., a diagonal line) is called a *secondary axis*. The *radius* of curvature of the lens is the radius of a sphere of which the refracting surface forms a part. Rays passing through the principal axis are not refracted, for the incident and the emergent ray strike each surface perpendicularly and the two surfaces at these points are parallel (see above). Moreover, rays in the secondary axes undergo only very slight refraction and the incident and emergent rays, though not quite in a continuous line, are parallel, for again the surfaces which they pierce are parallel. The point where the principal axis is intersected by the secondary axes is termed the *optical center* or *nodal point* of the lens. In a biconvex lens with symmetrical surfaces the actual and the optical centers coincide, but in other biconvex lenses the nodal point may be situated nearer to one or other surface.

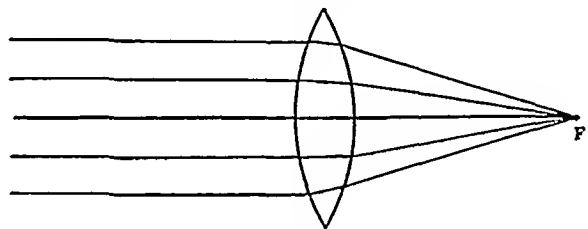


FIG 75 11 See text.

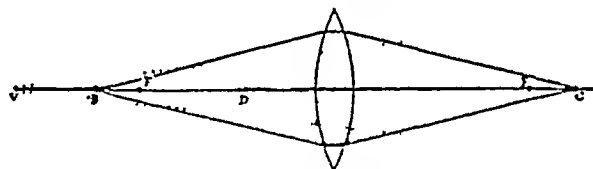


FIG 75 12 See text.

Light rays from a *distant object*, i.e., from infinity, are parallel (fig 75 11). The point F where parallel rays meet after refraction is called the *principal focus*. The distance of this point from the lens is called the *focal length* or *focal distance of the lens*. The rays of a light placed at the principal focus traverse the same path but in the opposite direction, and emerge as parallel rays.

A *near object* emits divergent rays. If the source of light (fig 75 12, B) is on the principal axis a little beyond the principal focus, an image is formed at a distance on the other side of the lens greater than its focal length (at C), when the object is placed in the second position (C) an image is formed at the first (B). These points, therefore, in respect to an object and its image, are interchangeable and are termed *conjugate foci*. If the object is situated at a distance exactly double the focal length of the lens the conjugate foci are at equal distances on the two sides of the lens. In all these instances a *real inverted image*, smaller than the object, is formed. If the source of light (D) lies between the lens and its principal focus, the rays, after passing through

the lens, are widely divergent. To the eye placed in the path of the emergent rays they appear to come from a point (V) at a greater distance behind the lens than the actual A *virtual, erect and enlarged image* is formed

Light rays in passing through a biconcave lens are diverged (fig 75 13), therefore a *true image* is not formed The eye in the path of the rays takes no account of refraction and the rays in consequence are projected backwards as straight lines which, meeting at a point on the other side of the lens, form a *virtual erect image, smaller* than the object.

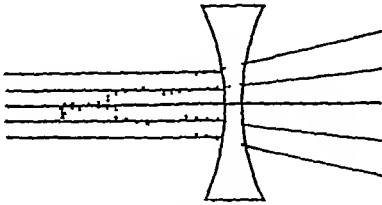


FIG 75 13 See text.

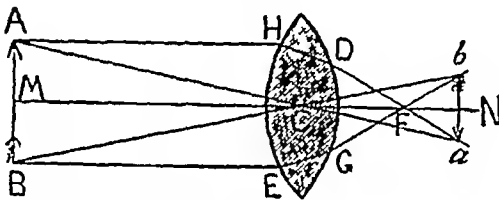


FIG 75 14 Formation of an inverted image by a biconvex lens

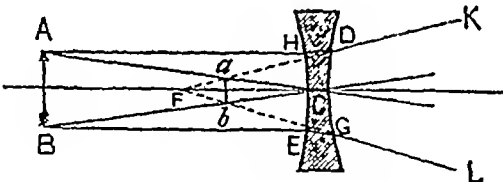


FIG 75 15 Formation of a virtual erect image by a concave lens

The formation of an image by a biconvex lens is shown in figure 75 14 AB is an object The rays AH and BE lying parallel to the axis MN meet and cross at the principal focus F The image of the point A will be at the intersection of the secondary axis AC with AHDF and that of B at the intersection of BC with BEGF Thus, a small inverted image is formed at ab The formation of an image by a concave lens is shown in figure 75 15 Rays AHDK and BEGL from points A and B of the object AB are diverged in passing through the lens and appear to come from the principal focus F, as shown by dotted lines The image of A is formed at a and of B at b, where these lines are intersected by the secondary axes

Cylindrical lenses (fig 75 17) have one plane surface, the other may be *convex* (*convex cylindrical lens*) or

concave (*concave cylindrical lens*) A convex cylindrical lens may be regarded as a section of a cylinder sliced down its long axis, its horizontal meridian is convex Rays transmitted through it at right angles to its vertical axis are converged as they would be by a convex spherical lens Light traversing its long axis is not refracted, the lens acting in this axis as a plate with parallel sides A lens of this type may be looked upon as an infinite series of prisms arranged base to base in tiers In the other type of cylindrical lens the horizontal meridian is concave, rays passing at right angles to the vertical axis are diverged

The diopter The converging or the diverging power of a lens depends upon the curvature of its surfaces (the greater the degree of curvature, the greater the refracting power) as well as upon the refractive index of the material of which it is composed The focal length of a lens varies inversely with the refractive power and is therefore a convenient measurement for expressing the strength of a lens The standard focal length is

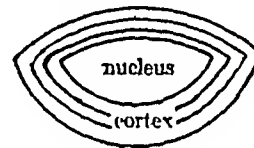


FIG 75 16 Showing structure of the crystalline lens (diagrammatic)

taken as 1 meter The refracting power is expressed as the reciprocal of the focal length (given in meters), the unit being the *diopter* (abbrev D) Thus, the strength of a lens with a focal length of 1 meter is 1 diopter, of one with a focal length of 2 meters, $\frac{1}{2}$ a diopter, of one having a focal length of $\frac{1}{4}$ a meter, 2 diopters, and so on The symbols + or - (+1 D, +2 D, -1 D, -2 D, etc) are used respectively for a converging and a diverging lens For example, if a concave lens has a refracting power of -1 D, a small virtual image of a distant object will be focussed 1 meter from the lens and on the same side as the object A convex lens of a corresponding power (+1 D) will bring parallel rays to a true focus 1 meter behind the lens The power of a cylindrical lens is expressed in a similar fashion

THE REFRACTING MEDIA OF THE EYE These are the cornea, the aqueous humor, the crystalline lens and the vitreous body The refractive indices are given in the following table

Cornea	1 37
Aqueous humor	1 33
Crystalline lens (whole)	1 42
Vitreous body	1 33

It will be noticed that the refractive indices of the cornea and of the aqueous humor and vitreous

are approximately the same, for practical purposes they may be taken as identical and the eye then taken as having two refracting surfaces, (a) the anterior surface of the cornea in contact with air, and (b) the lens surrounded by a common medium in so far as refraction is concerned. The greatest refraction occurs at the corneal surface (42 diopters), of less importance is refraction at the surfaces of the lens (19 diopters with accommodation relaxed and 36 diopters in full accommodation). The whole eye has a refracting power of between approximately 60 and 65 diopters. The value in the table above for the refractive index of the lens is calculated from the refractive power of the lens as a whole, as though it were a homogenous structure, but such is not the case. On the contrary, the lens consists of an almost spherical *nucleus* with a high refractivity (1.41) surrounded by a zone called the *cortex* of lower optical density (1.38). The surrounding cortex is composed of a series of concave meniscus lens, as shown in fig 75 16. Thus an image formed by the anterior part of the cortex (B) is focussed by A and the second image in turn by C. The peculiar structure of the lens accounts for the paradox that the mean value of the refractive indices of its parts (1.39) is less than the refracting power of the whole. Several important advantages are derived from this peculiar structure of the crystalline lens, it diminishes spherical and chromatic aberration (p 1160), tends to prevent the scattering of light within the eye and enhances the power of the lens to alter its converging power during accommodation (p 1152).

THE CONSTANTS OF THE EYE Knowing the curvatures of the refractive surfaces of the eye and the distances between them as well as the refractive indices of the media, the path taken by the rays of light can be determined and the image constructed. The values of the *constants of the eye* are given in the following table

	mm
Position of anterior surface of cornea	0
Position of posterior surface of cornea	0.6
Position of anterior surface of lens	3.6
Position of posterior surface of lens	7.2
Position of retina	24.1
Radius of anterior surface of cornea	7.7
Radius of posterior surface of cornea	6.8
Radius of anterior surface of lens (distant vision)	12.0
Radius of posterior surface of lens	6.0

The refractive indices of the media have been given above

Construction of the image from the foregoing data is a very laborious proceeding. To start with, the image formed by the first refracting surface is constructed, this image now serves as the object for the next refracting surface, the second image in turn is the object of the third surface and so on. The matter is very much simplified, however, by constructing the *diagrammatic* or *schematic* eye by the application of the theorem of Gauss. This states that every optical system composed of spherical surfaces with their centers on the principal axis has three pairs of cardinal points. These are, two *principal points* (K and K', fig 75 18), an *anterior* and a *posterior focal point* (φ and φ') and two *nodal points* (N and N').

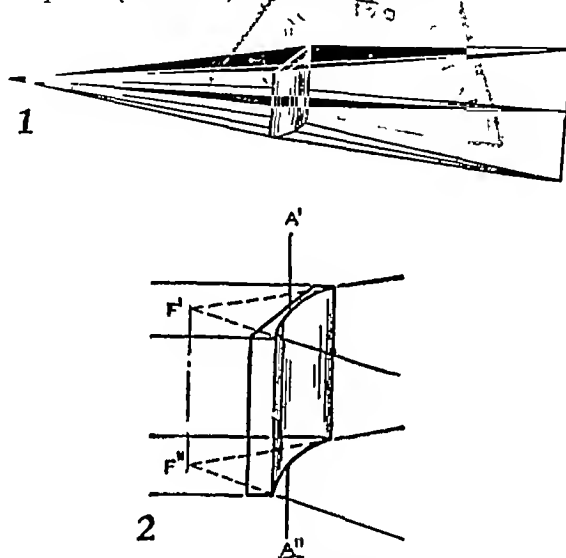


FIG 75 17 1, refraction by a convex cylinder. A point of light is brought to a focus as a line after refraction through a cylinder. 2, refraction of light by a concave cylinder. Rays of light striking the cylinder perpendicularly to the axis A'A" are diverged, and appear to be brought to a virtual focal line F'F" (After Duke-Elder)

The *first* and *second principal points* lie close together in the anterior chamber 2 mm behind the cornea. Planes passing through the principal points and perpendicular to the axis are termed the *first* and *second principal planes*, an object in the first principal plane forms an erect *full-sized image* in the second and vice versa. The first and second principal points correspond, therefore, to the conjugate foci of a single lens.

The anterior focal point (φ) is situated 15.7 mm in front of the cornea. Rays from this point in the axis would, after passing through the system, emerge as parallel rays. Parallel rays entering the system are focussed at φ' which is situated on the retina. φ and φ' therefore correspond to the principal foci of a single lens.

The nodal points (K and K') also lie close together on the axis and near the posterior surface of the lens.

They correspond to the optical center of a single lens. Rays passing through the nodal points are not refracted. A ray entering the system and passing through K appears to come from K' and emerges along a line parallel to that along which it entered. The following table for the schematic eye gives the distances of the six cardinal points from the anterior surface of the cornea.

	mm
Anterior surface of cornea	0
First principal point, H	1.7
Second principal point, H'	2.0
First nodal point, K	7.0
Second nodal point, K'	7.3
Posterior focal point, ϕ'	24.1
Anterior focal point, ϕ	15.7

The two nodal points lie so close together that no

THE FORMATION OF THE IMAGE ON THE RETINA. Knowing the foregoing measurements, the paths taken by the light rays can be drawn and a construction of the image upon the retina readily carried out. The formation of the retinal image is illustrated in fig. 75.19. The large arrow A-B represents an object emitting divergent rays which are converged and brought to a focus to form the image represented by the small arrow a-b. The retinal image, it will be observed, is smaller than the object and inverted. For the sake of simplicity only a few rays are shown, two from a point at either end of the object and one from its center, but of course the surface of an object consists of an infinite number of points, each of which emits

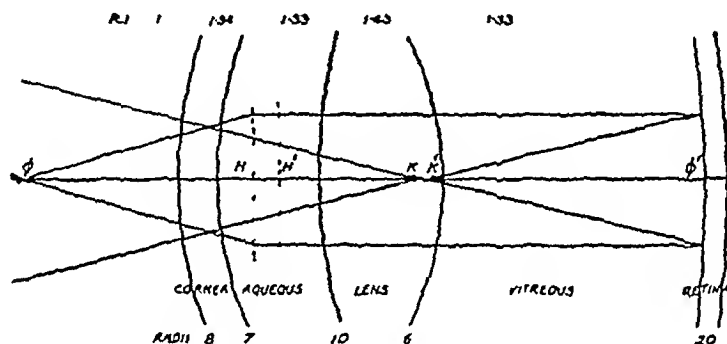


FIG. 75.18 The cardinal points of the eye. ϕ , The anterior focus, 15.7 mm in front of the cornea. ϕ' , The posterior principal focus, 24.13 mm behind the cornea—that is, upon the retina. H , H' , the principal points, in the anterior chamber. K , K' , the nodal points, in the posterior part of the lens. (After Duke-Elder)

significant error is entailed if they are taken as one, the same may be said for the principal points. Thus the compound optical system of the eye can be simplified to the so-called *reduced schematic eye* of Listing. This has a single ideally refracting surface situated in the anterior chamber 1.35 mm behind the cornea and with a radius of 5.7 mm. The nodal point or optical center of the reduced eye lies 7.08 mm, the principal point 2.3 mm and the posterior focal point 24.13 mm behind the anterior corneal surface. The anterior focal point is 15.7 mm in front of the cornea. The distance of the nodal point from the retina, i.e., the focal length of the eye, is $(24.13 - 7.08 =) 17.05$ mm. The refracting power is therefore $(1000/17.05 =) 58.65$ diopters.

By means of an X-ray beam, which is not refracted, projected into the eye Goldman and Hagen have measured the length of the globe in the living human subject. The value obtained (23.4 mm) agrees closely with that of the schematic eye. The value for the total refractive power of the normal human eye, as determined by these observers, is also in close agreement, namely, 59.22 diopters.

divergent rays. One of each pair of rays in the figure (solid line) passes unrefracted through a secondary axis (i.e., through the nodal point N), the ray from the upper part of the object to the lower part of the retina and vice versa. The other ray of each pair undergoes refraction and meets the corresponding unrefracted ray. Similarly, rays from a point on one side of the object will fall upon the retina as a point of light in the opposite part of the image. Thus, it is seen how the image on the retina becomes inverted. Of course we see objects in their true position. Re-inversion is a cerebral function developed, probably, through the association of visual sensations with those of touch. The process itself is essentially psychological and beyond our ability to analyze.

The idea that the retinal image is inverted was at first difficult to believe. Kepler (1604) inferred from his optical studies that this must be so, but it was Scheiner (1625) who furnished the proof by observing the back of an excised eye from which

the sclerotic and choroid coats had been removed. The inverted image of an object was clearly visible upon the translucent retina. The inversion of the retinal image may also be demonstrated during life in persons of blonde complexion because their choroid contains little pigment. The subject is examined in a darkened room, the eye being turned towards a lighted candle placed well to the temporal side. An inverted image of the flame may be seen showing through the inner side of the wall of the globe.

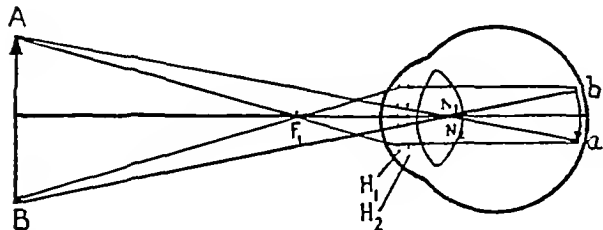


FIG 75.19 Illustrating the inversion of the retinal image (schematic eye), F_1 , first focal point, H_1 and H_2 , first and second principal points, M_1 and N_2 , first and second nodal points, AB , object, ab , image

The size of the retinal image is dependent upon the angle $a N b$, (fig 75.19) this—the angle subtended at the nodal point of the eye by an object in the visual field—is called the *visual angle*. The size of the image can be calculated if one knows the distance of the nodal point of the eye from the object and from the retina, and the size of the object. Thus in the figure AB is the object and ab its image. The triangle ANB and aNb being symmetrical, then

$$a b A B = b N B N$$

$$a b = A B \times b N / B N$$

$$b N = 17.05 \text{ mm}$$

$$z = 17.05 \times O/d$$

z is the size of the image, O the size of the object and d is distance from the nodal point of the eye (17.05 = distance of retina from nodal point, see p 1146)

THE OPHTHALMOSCOPIC EXAMINATION OF THE EYE Under ordinary circumstances we cannot see within the eye of another person, because only a limited quantity of light enters his eye through the relatively small pupillary aperture. It is like trying to look through a small window into a darkened room. Furthermore, of the light which enters the eye, a large part is absorbed by the pigment layer of the retina. Even when a light is brought close to the eye under observation, one is unable to see the retina, because the pencil of parallel rays which emerge do not enter the examiner's pupil unless his eye is directly in its path, and when he attempts to bring his eye into the

proper position either his head comes between the light and the subject's eye or the light (if between himself and the subject) dazzles his sight.

These difficulties are overcome by means of the ophthalmoscope. This instrument consists of a small mirror with a central perforation through which the observer views the subject's eye. Light furnished by an electric lamp placed above the head of the patient is reflected from the mirror and the pencil of rays emerging from the subject's eye passes through the aperture of the mirror to the examiner's eye (fig 75.23).

The invention of the ophthalmoscope is commonly attributed to Helmholtz (1851) but a crude device based upon the same principle was used by Babbage two years previously. The instrument consists of two mirrors, each with a central aperture. One mirror is plane, the other concave, they are pivoted so that one or other can be rotated into position as required. The obverse side of the instrument holds a series of small lenses ranging from +30 D to -30 D in strength, any one of which can be moved into position over the central aperture. The examiner seats himself 1 meter from the subject and looking through the peephole of the plane mirror reflects a beam of light into the eye. If the observed eye is emmetropic (i.e. of normal refraction) a uniform red glow—the *red reflex of the fundus*—caused by reflection from the retina is seen lighting up the pupil.² No detail, e.g., optic disc or retinal vessels, is visible. This is because the rays emerge from the subject's eye as two diverging pencils of parallel rays which come from opposite points of his retina. In order for an image of the subject's fundus to be formed upon the examiner's retina, rays from both pencils must enter his eye simultaneously. At a distance of 1 meter this is impossible, but can be effected if the observer brings his eye quite close to the patient's eye (see figure 75.20). When the subject's eye is hypermetropic the rays composing the emergent pencils are *divergent*, it is therefore possible at a distance of 1 meter for rays from both to enter the observer's eye. The rays appear to meet behind the subject's eyes where a *virtual erect* image is formed (fig 75.21). When the observer moves his head the image moves in the *same* direction. In myopia, the emergent rays *converge*, they meet and cross in front of the eye. Again, diverging rays from both bundles enter the observer's eye and form a *real inverted* image in front of the subject's eye (fig 75.22),

² This is similar to the red glow seen in the eyes of certain animals, especially of the cat and other carnivora, which is due to the presence of a highly polished area at the fundus—the *tapetum lucidum*—which acts as a concave mirror to throw back a large proportion of the light which enters the eye. This structure lies behind the retina, the pigment layer being absent over it. It is composed of several layers of flattened endotheal cells overlain with doubly refracting crystals.

it moves in a direction *opposite* to that of the examiner's head.

After this preliminary examination one or other of two methods of ophthalmoscopy—the indirect or the direct—may be employed. In the *indirect method* a separate biconvex lens with a focal length of 7.5 cm (about 13 D) is held in front of the eye of the examinee. The observer seats himself 1 meter away and holds the lens in the path of the beam from the mirror and a short

The rays reflected from innumerable points of the patient's retina and after refraction emerge from his eye to form an image, the position of which differs according to the refractive state of the eye. The rays emerging from the emmetropic (normal) eye are parallel (fig 75.24A). They are converged by the hand lens and form a real, magnified and inverted image in the air

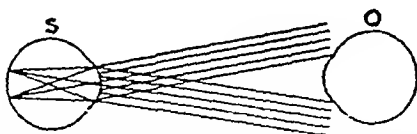


FIG 75.20 Showing two pencils of parallel rays as they emerge from the normal eye S, subject, O, observer. Vertical dotted line indicates the position in which both pencils of rays will enter the observer's eye (After Duke Elder)

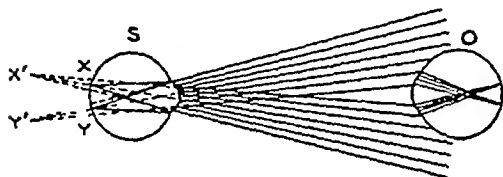


FIG 75.21 Emergent rays from hypermetropic eye. A virtual erect image is formed at $x'y'$ (After Duke Elder)

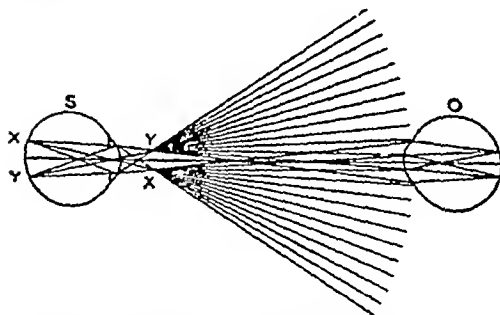


FIG 75.22 Emergent rays from myopic eye which cross at $x'y'$ and then diverge, a real inverted image is formed at $x'y'$ (After Duke Elder)

distance in front of the eye under observation. The subject turns his eye a little inwards in order to bring the optic disc into view, being directed, for example, to look at the observer's left ear if his left eye is being examined (fig 75.23). The diverging rays from the light source are reflected from the mirror and converged upon the hand lens which converges them to the eye. They are more sharply converged by the refracting media and come to a focus in the vitreous. The retina is diffusely illuminated. By moving the hand lens toward or from the patient's eye his retina is brought into focus and the optic disc (Plate III) is clearly seen

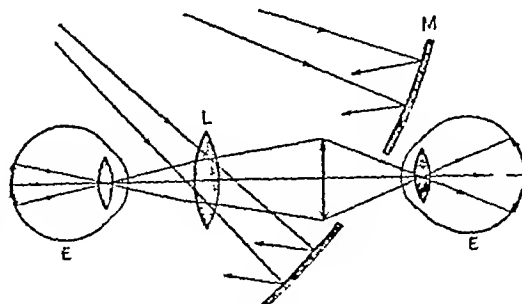


FIG 75.23 Showing the course taken by the light rays in indirect ophthalmoscopy. The mirror, M, is represented as several times actual size. E, observer's eye, E', subject's eye, L, hand lens. The arrow represents a real inverted image formed between the lens and the observer

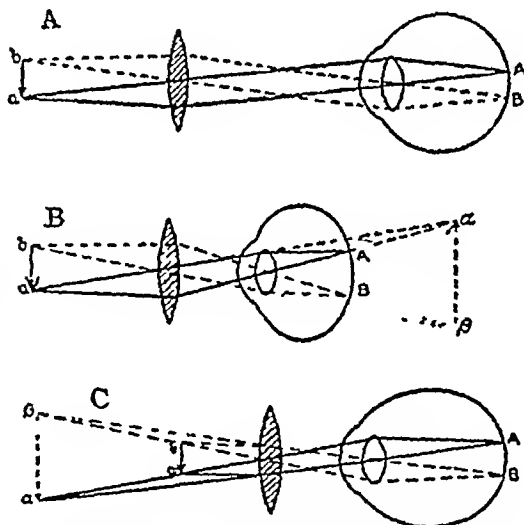
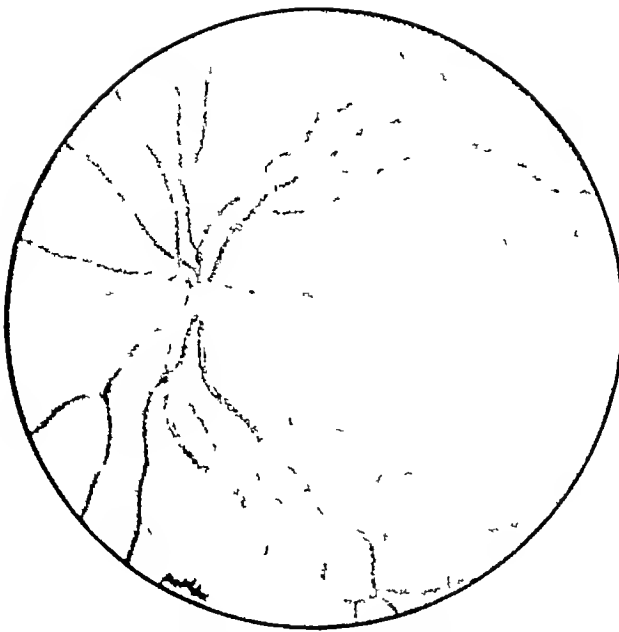


FIG 75.24 Path of the light from the eye in the indirect method of ophthalmoscopy in emmetropia, hypermetropia and myopia (After Duke-Elder, *Text Book of Ophthalmology*)

between the hand lens and the examiner. This is focussed on the examiner's retina. In *hypermetropia* the rays issuing from the patient's eye are divergent, and appear to come from a magnified erect image behind the eye. The hand lens converts this into a small, inverted image in front of its principal focus (ab , fig 75.24B). In *myopia* the emergent rays converge and form a large inverted real image ($\alpha\beta$), this is converted

PLATE III



Normal Human Fundus (from May's *Manual of Diseases of the Eye* by permission of the author)

by the lens into a smaller image (ab, fig 75 24C) between it and its principal focus

In the *direct method* a field of smaller area but of higher magnification is seen. The light is placed behind the patient's head and to one or other side, depending upon the eye to be examined. The examiner brings the instrument as close as possible to the subject's eye, no hand lens being interposed. He views the subject's fundus with his own right eye, and the subject's left fundus with his left eye. In this position rays from both emergent pencils enter the eye of the observer, and, therefore, except when the observed eye is highly myopic, the details of the fundus can be seen. The image is always erect. The small concave mirror, which is set obliquely in the instrument, is rotated to the proper angle to throw a reflected beam through the pupil. The light, after coming to a focus on the retina, emerges from the emmetropic eye as parallel rays which are brought to a focus upon the observer's retina. If, however, the subject's eye is hypermetropic (as described above) the emergent rays will be divergent and can only be focussed by the examiner if he accommodates his eye or interposes a convex lens, if the observed eye is myopic the reflected rays are convergent, the examiner must then use a concave lens in order to focus the image upon his own retina. By the use of a concave or a convex lens which will just bring the image of the fundus into clear focus, and the refractive power of the lens required being known, the refractive error of the patient can be measured.

RETINOSCOPY, SKIASCOPY or the Shadow test This is a reliable objective method for determining the refraction of the eye. It is of special value in children and others for whom the reading of test type is impracticable, and for the detection of malingerers. The method depends upon the fact discovered by Bowman (1859) that the direction of the rays emerging from the eye varies with its state of refraction. Caignet (1873) elaborated upon this discovery and brought the method into general use in the study of refractive errors. When a lighted candle is held in front of the eye, its rays undergo refraction and illuminate an area upon the retina. Since the retinal image is inverted, a movement of the candle to one or the other side, up or down, causes the illuminated area to move in the opposite direction.³ When the rays are reflected from a plane mirror they form a virtual image behind the reflecting surface. Tilting the mirror to one or the other side causes this image to move in the opposite direction. Consequently, when a reflected light is thrown into the eye by means of a plane mirror, owing to a double reversal of the movement taking place (i.e., a movement of the mirror image which now serves as the luminous object in one direction and of the retinal image in the other), a change in the inclination of the mirror causes the illumina-

nated area upon the retina to move in the *same* direction as the tilt. This movement of the illuminated area on the *retina* in relation to the tilt of the mirror is the same whatever the refractive state of the eye. But the rays from the illuminated area on the retina are directed back through the eye and after undergoing refraction emerge through the pupil, the direction of the movement of the illuminated area as it appears to an observer depends upon the direction of the emergent rays, that is, upon the state of refraction of the eye. The eye is examined with the ophthalmoscope at a distance of 1 meter, in a dark room, a mydriatic, e.g., atropine or homatropine is usually employed.

If the subject is *hypermetropic* or *myopic* a pink glow fills his pupil except for a dark semilunar shadow on one side (see fig 75 25). In the *emmetropic* eye light reflected

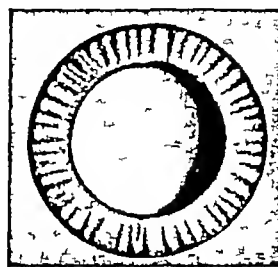


FIG 75 25 Retinoscopy shadow (After Duke-Elder)

from the illuminated area on the retina issue from the pupil as diverging pencils of *parallel* rays (p 1148). The rays meet and cross at the *far point* (punctum remotum), which for the normal eye is at infinity, that is, behind the observer. In a certain position of the mirror one or other of the pencils of rays (see fig 75 26) enters the observer's eye. The subject's eye appears filled with a pink glare bounded on one side by a very faint shadow, when the mirror is tilted so that the interval between the pencils is in line with the observer's pupil, the subject's pupil is dark. Thus, in emmetropia the subject's pupil is either nearly uniformly bright or entirely dark. A bright area and a pronounced shadow are not seen together. In *myopia* the rays leaving the eye are convergent (fig 75.26). If the myopia is *greater* than 1 D they meet and cross at the far point which lies somewhere between the subject and the observer. Therefore, when the mirror is tilted, up or down, or to one or other side, the shadow moves *against* it, i.e., in the opposite direction. If the myopia is *less* than 1 D the far point is behind the observer and, as in emmetropia, only a faint shadow is seen, it moves *with* the mirror. If the myopia is 1 D the far point is at the surface of the observer's eye and the subject's pupil appears, according to the tilt of the mirror, either completely dark or completely bright, that is, without any shadow. In *hypermetropia* the emergent rays are divergent and the far point is virtual, the rays appear to come from a point behind the mirror and therefore

³ If a concave mirror were used, the image being formed in front of the mirror, would move in the same direction as the tilt of the mirror.

do not meet between the subject and observer. The shadow moves *with* the mirror.

In applying these facts to the correction of refractive errors, the movement of the shadow is noted and convex or concave glasses, according to whether the eye is hypermetropic or myopic, respectively, are placed in front of the subject's eye, until the pupil, as in a myopia of 1 D, appears uniformly bright or dark. This state when the shadow disappears is called the *point of reversal*. If the refractive error is even slightly overcorrected a shadow appears which moves in a direction opposite to that of the original movement. Now, when the point of reversal is reached, the far point of the

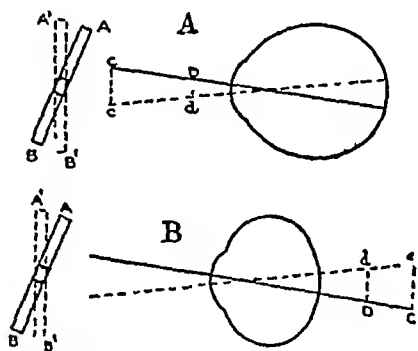


FIG 75.26 Illustrating the movements of the shadow in retinoscopy. A, in *myopia*. When the mirror is in the position AB, the image is at D in the highly myopic eye, and at C in the less myopic eye. When the mirror is in the position A'B', the respective images are in the positions d and c. Since Dd is less than Cc, the lower the degree of myopia, the larger the excursion and the quicker the movement of the shadow. B, in *hypermetropia*. When the mirror is in the position AB, the image is at D, in the highly hypermetropic eye and at C in the less hypermetropic eye. When the mirror is in the position A'B' the respective images are at d and c. (After Duke-Elder, *The Practice of Refraction*)

subject's eye is at the surface of the observer's eye, i.e., 1 meter away. The subject therefore, as mentioned above, has still a myopia of 1 diopter. This must be taken into account in calculating the refractive power of the correcting lenses required. If the eye is hypermetropic and, say +4 D brings it to the point of reversal, then a +3 D lens will be sufficient for correction. If the eye is myopic -1 D must be added to the refraction which was required to bring the eye to the point of reversal.

Papilledema (choked disc, optic neuritis) and optic atrophy. In conditions accompanied by high intracranial pressure, e.g., brain tumor, hydrocephalus and uremia, the optic disc loses its natural translucency and becomes reddened and swollen. The central vein is engorged and tortuous and the venules and capillaries dilated. Small hemorrhages may be seen. The swelling of the disc is due to the transudation of fluid from

the engorged vessels and its collection in the anterior layers of the lamina cribrosa and between the nerve fibers. The physiological cup becomes gradually filled up, and may eventually be elevated above the general level of the surrounding retina. The circumference of the disc ill-defined, appearing blurred or "woolly." The disc is enlarged and its lateral spread causes the retina to be thrown into folds or ridges. In the older terminology these changes were referred to as choked disc, optic neuritis or papillitis. The condition is now called *papilledema*. The separation and stretching of the nerve fibers, their compression where they penetrate the lamina cribrosa and the overgrowth of glial tissue set up by the presence of the edema fluid leads to nerve atrophy—*secondary optic atrophy*. The disc in this condition is a grayish or dead white, due to the obliteration of capillary vessels by the overgrowth of glial tissue. The optic cup is deepened as a result of the degenerated nerve fibers. The outline of the pale disc is clearly defined against the surrounding retina. The retinal veins are engorged but the arteries are narrower than normal.

It is now widely accepted that mechanical factors, e.g., changes in intracranial or intra-ocular pressure, are mainly concerned in the production of papilledema. The optic nerve, it will be recalled, is invested by prolongations of the cerebral meninges—the *pia*, *arachnoid* and *dura*. The spaces between these three layers of the nerve sheath—the *subarachnoid* and *subdural spaces*—are continuous, as was first shown by Schwalbe (1870) with the corresponding spaces within the cranium. The *dura* and *arachnoid* of the nerve sheath, however, are in close opposition, there being only a potential subdural space in this situation. But the intracranial subarachnoid space is in free communication with the corresponding space of the nerve sheath and the intracranial pressure is transmitted through the cerebrospinal fluid to the intravaginal space right up to the lamina cribrosa. The central vein of the retina with its companion artery makes an almost right angled bend as it leaves the optic nerve and, crossing the intravaginal space, pierces the arachnoid and *dura* a short distance behind the eyeball (fig 75.27). Any marked rise in intracranial pressure distends the sheath, for the *dura* in this situation differs from that within the cranium in being unsupported by bone (Macdonald). The elevated pressure thus transmitted to the sheath tends to compress the vein in its course across the intravaginal space and to impede the venous return, as well as to block the lymph channels situated in the adventitia of the central vessels. Little or no interference is offered to the blood flow in the artery owing to its more resistant wall and to the higher arterial blood pressure. The venous pressure rises and the intraocular part of the vein and its branches become engorged since they are beyond the influence of the intracranial pressure.

Papilledema is therefore comparable with an edema

which may occur in almost any situation as a result of obstruction of the venous and lymphatic channels while the arteries remain pervious. No evidence that papilledema is inflammatory in nature can be found upon histological examination (Holmes and Paton). The theory that high intracranial pressure causes edema of the disc by raising the pressure in the cavernous sinus (which receives blood from the central vein either directly or through the superior ophthalmic vein) is controverted by the following facts: (a) The central vein of the retina after piercing the dura anastomoses within the orbit with one or other of the ophthalmic veins (usually the superior) and through it with the pterygoid venous plexus and the facial vein. Thus an alternative system of channels is provided for the drainage of blood from the retina. (b) Ophthalmic changes are not seen as a rule in thrombosis of the cavernous sinus.

It has been suggested by Macdonald that cases of retinal detachment which sometimes follow a powerful straining effort (e.g., lifting a heavy weight) may be due to an acute retinal edema caused by the sharp rise in intracranial pressure which such acts induce. In his view the distension of the nerve sheath by cerebrospinal fluid causes the central vein to be bent more acutely at the point where it leaves the nerve and that the venous congestion is caused in this way rather than by pressure upon the vessel in its subarachnoid course.

Atrophic changes in the disc occurring unpreceded by papilledema are referred to as *primary optic atrophy*. The main causes of primary optic atrophy are pressure upon the nerve within the cranium as by a tumor, certain nervous diseases (tabes, general paralysis of the insane and disseminated sclerosis) and toxic substances, e.g., wood alcohol, quinine, lead and salicylic compounds. Since the optic nerve fibers have their cell bodies in the retina (ganglion cell layer) the atrophy of the section between the point of pressure and the retina is in the nature of a retrograde degeneration (p. 912). As already mentioned on (p. 913), the optic nerve fiber is devoid of a neurilemma, regeneration therefore never occurs. In tabes and general paralysis of the insane the atrophy is probably the result of a syphilitic meningitis which affects the nerve secondarily. The toxic substances mentioned exert their action apparently directly upon the ganglion cells of the retina.

BIOMICROSCOPY—THE SLIT LAMP By means of this instrument an intense narrow beam of light is thrown obliquely into the eye and a small section of the cornea, anterior chamber, iris, lens or anterior part of the vitreous observed stereoscopically through a binocular microscope. The beam passes through an adjustable slit which when narrowed to minimum width concentrates the beam of light upon an area as small as 0.05 mm in diameter. The illuminated section is in the form of a prism, or more correctly of a parallelepiped. The tissues within this section are magnified some 25 diameters. Under this method of examination ocular

structures which ordinarily appear homogeneous show a definite pattern, and any abnormality is readily recognized by one familiar with the appearance in health. For example, the anterior and posterior epithelial layers of the cornea are seen as bright lines bounding the less luminous substantia propria, and any pathological condition, e.g., erosions, small opacities, keratitis, etc., are easily detected. The laminated structure of the lens is clearly revealed. The central nucleus and the cortical layers are marked by luminous boundaries, the lens sutures appear as darker lines.

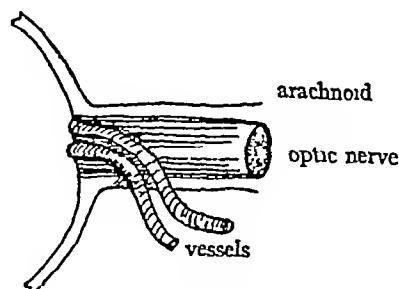


FIG 75 27 Description in text.

ENTOPTIC PHENOMENA Visual sensations may arise from images of objects situated within the eye itself. The most familiar of these are the *muscae volitantes* (L. flying flies) which are seen as faint specks projected a short distance in front of the eye. They are due to shadows cast upon the retina by small semi-opaque particles in the vitreous body, such as epithelial cells, small coagula or embryonic rudiments. Particles lying close to the retina are most likely to give rise to these sensations. Lying behind the axis of rotation of the globe such particles cast shadows which move downwards over the retina when the eye is turned upwards but, since the direction of any movement on the retina is reversed in consciousness, they appear to move upwards. This upward movement is followed by a slower downward movement. When any attempt is made to fix one's sight upon the specks they dart away, from which fact their name was derived. The movements of shadows caused by particles in the anterior chamber or in any part of the eye in front of its axis of rotation would, of course, be in the reverse direction.

Ordinarily the retinal vessels which, it will be recalled, lie outside the fovea and superficial to the retinal layers, are not perceived. One probable reason for this is that we have come through habit to ignore them, but other factors of a more physiological or anatomical nature have been suggested. Helmholtz, for example, thought that the sensitivity of the retina underlying the vessels was

greater than elsewhere, so that the light reaching the retina through them, though reduced in intensity, caused as great an effect as light falling in unshielded regions. At any rate, when light is thrown into the eye at such an angle (e.g., obliquely through the sclerotic) that the shadows of the vessels fall upon a retinal region unaccustomed to receive them, they become visible. If, while his eye is being illuminated in this way in a dark room, the subject looks towards a wall, and the light (e.g., a candle) given a slow circular movement, he sees a highly magnified image projected against the uniform surface. The vessels appear as an intricate branching pattern against a bright ground and are known, after their discoverer, as *Purkinje figures*. A method which anyone can employ himself to make the extrafoveal capillaries visible is to look at a uniformly bright surface through a pin-hole in a card held close to the eye while he oscillates the opening quickly (about once per second) from side to side and thereby shifts the shadows from point to point in the retina. The capillaries may also be observed by looking through a microscope from which the objective (but not the eyepiece) has been removed and moving the head rapidly from side to side.

Mueller (1855) made use of Purkinje's observation (1811) to prove that the light sensitive elements are the rods and cones. Moving the source of light causes the images to change their positions on the screen. If we measure the distance of this shift (fig. 75.28, A-B) and the distance between the two positions of the light (a-b) then, knowing the distance of the nodal point of the schematic eye (N) from the retina and from the screen, the position of the shadow ($\alpha-\beta$) relative to the vessels (v), i.e., the distance $\alpha-v$, can be calculated. This was found to be from 0.17 to 0.36 mm, which is by actual measurement the approximate distance of the vessels in front of the rod and cone layer.

The corpuscles moving in the retinal capillaries can be observed if the eye is directed to a uniformly illuminated surface. The best way to perform this experiment is to look at the sky through a dense blue-violet glass plate. The blood cells then appear projected upon the plate. It is actually possible to calculate the speed of the corpuscles from the distance between their positions at the beginning and end of a given time interval. Vierordt (1873) was the first to make such an estimation. Knowing the distance of the nodal point of the eye from the retina and from the glass screen, the magnification of the travelled distance can be determined (p. 1147).

Other entoptic phenomena which should be briefly mentioned are the colored halos seen around bright lights especially in dark surroundings. The halos

consist of a series of concentric rings of rainbow colors—from blue to red from within out. Actually they are diffraction spectra, due to the structures of the eye acting as diffraction gratings. There are two kinds of halo—*lenticular* and *corneal*. The first is attributed to the radial fibers of the lens and is the larger. The corneal type is believed to be due to the epithelial and endothelial cell layers of the cornea.

THE ACCOMMODATION OF THE EYE ANATOMICAL SKETCH

Before giving an account of the physiological mechanism of accommodation, the structures concerned, e.g., the ciliary body and the crystalline lens will be briefly described.

THE CILIARY BODY When the interior of the anterior half of the eyeball is exposed by a transection through its equator, a transparent disc—the *crystalline lens*—is seen occupying the center of the bowl shaped structure (fig. 75.29). On the wall of the globe some

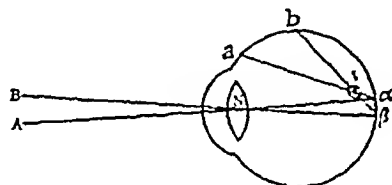


FIG. 75.28 Description in text.

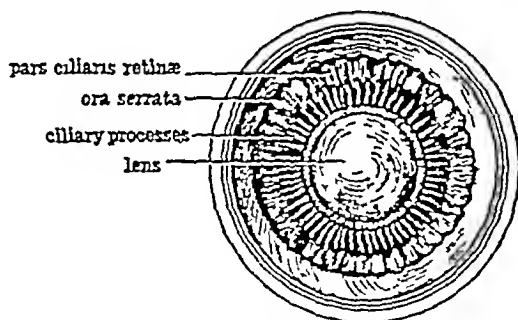


FIG. 75.29 Showing interior of anterior half of the eyeball.

distance behind the circumference of the lens lies the dendate border of the retina proper known as the *ora serrata* (p. 1111). The ciliary body is a circular zone of tissue extending forwards from the ora serrata to a short distance from the circumference of the lens. It is covered on its inner aspect by the pigment layer of the retina, which we have seen (ch. 73) is continued forwards (as the *pars ciliaris retinae*) from the point of termination of the neural layers. The ciliary body consists of three parts, the orbiculus ciliaris, the ciliary processes and the ciliary muscle. The *orbiculus ciliaris* immediately adjoins the choroid, of which it may be considered the direct continuation. It is a band about

4 mm broad encircling the eyeball and presents on its inner aspect a number of radially arranged ridges. The *ciliary processes* appear as some seventy triangular elevations on the inner aspect of the ciliary body, they project towards the axis of the eye and form a series of radial fringes (*corona ciliaris*) which completely encircle the equator of the lens, but are separated from it by a very short interval. The great bulk of the ciliary body is composed of the ciliary processes and the ciliary muscle. The fibers of the *ciliary muscle* are arranged in two sets, an outer *meridional* and an inner *circular*. The meridional fibers arise from the

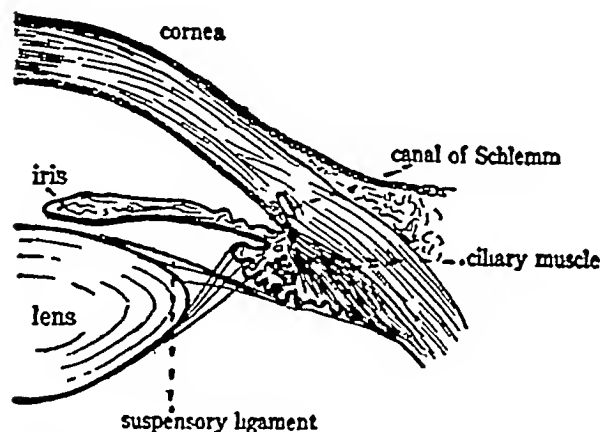


FIG 75.30 Showing structures in the region of the angle of the iris



FIG 75.31 Showing arrangement of circular fibers of the ciliary muscle (After Fincham)

scleral spur (p 1165), they run backward to be attached to the ciliary processes and to the orbiculus, and through the latter to the choroid. The circular fibers are fewer, in meridional sections of the globe they appear as a small triangular bundle of cross-sectioned fibers lying behind the angle of the iris (fig 75.30). As a matter of fact, these fibers are not uniformly circular in direction but take different courses and, interlacing with one another, form a reticulated ring-shaped band (fig 75.31). Taken as a whole this part of the ciliary muscle constitutes a sphincter (*sphincter muscle of Mueller*) situated in front and to the outer side of the ciliary processes. The fibers composing the margin of the central opening are mainly circular and are attached to a band of elastic tissue situated at the angle of the iris, the outer circumference of the sphincter is connected to elastic fibers which are continuous with similar fibers of the choroid. The muscle is thus anchored by two elastic attachments.

THE CRYSTALLINE LENS is a transparent, biconvex, circular structure about 11 mm in diameter and between 3.6 and 3.9 mm thick at the center. It is situated with the center of its anterior surface coinciding with the center of the pupil, the pupillary margin lies in contact with this surface. The center of the anterior surface is termed the *anterior pole* of the lens, the center of its posterior surface, the *posterior pole*. An imaginary line joining the poles is called the *principal axis*. The two surfaces meet at the circumference in a rounded edge termed the *equator*. The posterior surface in the young adult is decidedly more convex than the anterior (see table, p 1145), this difference diminishes somewhat with age. The lens is enclosed in a structureless, highly elastic capsule. The latter is not of uniform thickness, being thinner over the posterior than over the anterior surface, and the part covering the central region of each

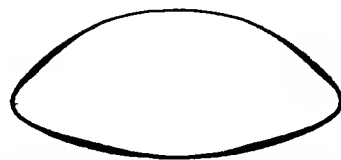


FIG 75.32 Showing the regional variations in thickness of the lens capsule (After Fincham)

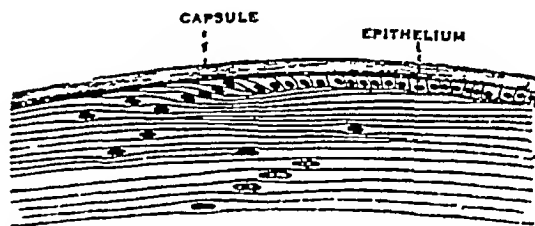


FIG 75.33 Meridional section of the lens (From Wolff after Poirier and Becker)

surface is thinner than the corresponding peripheral parts. The values in ascending order of thicknesses are, center of posterior surface (av 2.2μ), peripheral part of posterior surface (av 13.7μ), center of anterior surface (av 15.7μ), peripheral part of anterior surface (av 18.2μ) (see fig 75.32). A single layer of columnar epithelial cells covers the anterior surface of the lens immediately beneath the homogeneous capsule, the latter is formed as a secretion of these cells. The substance of the lens consists of a series of ribbon-like fibers which arise from the region of the equator and are actually greatly elongated epithelial cells (fig 75.33). By careful examination of the lens, from the more central part of the anterior surface to the region of the equator the gradual transition of the columnar cells into the attenuated cells of the lens substance can be traced. The fibers proceed from the equator towards the lens center and, abutting against fibers coming from other segments of the periphery, fuse along well-defined lines—the *lens sutures*. These are seen in the adult lens as a series of faint irregular striae radiating

from the center to form what is known as the *lens star*. On section, the lens shows a series of concentric laminae with a nucleus of extreme convexity and high refractive index, and a less refractive cortex. The optical advantages of this construction have been pointed out (p 1145). The nucleus is also of much firmer consistency than the cortex which is relatively soft and pliable.

THE MECHANISM OF ACCOMMODATION

The interval between the ciliary processes and the equator of the lens is occupied by a circular membranous band, this is the anterior part of the *zonula ciliaris* (*zonula of Zinn*). The precise origin of the fibers of the zonula is disputed, but they appear to arise as a system

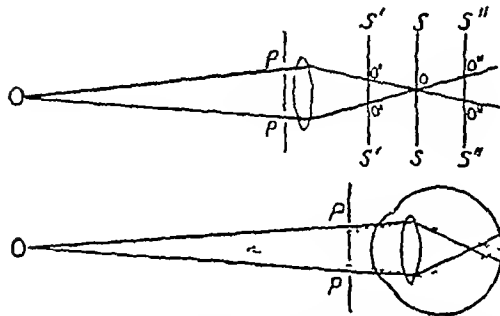


FIG 75.34 Scheiner's experiment. Upper cut, O, position of needle, PP, pinholes in card, SS, screen with image of needle focussed at o. When the screen is moved forward or backward (S'S' or S''S'') the two pencils of light are intercepted before they have reached a focus in the first instance, and after they have met and crossed in the second. Two images are formed at o' and o' or at o'' and o''. In the case of the eye (lower cut), any difference in its focus must be due to changes in the lens, since the retina is stationary. When the needle is viewed with one eye through the peep holes, and brought into focus one image is seen, but two blurred images appear if the eye is focussed upon a nearer or a more distant object. If when the eye looks at an object nearer than the needle, and the needle is moved a little nearer to the eye, the two images approach one another, when the needle is moved farther away they become more separated. When the eye is accommodated for far vision changes in the distance of the needle cause converse movements of the images, becoming farther apart as the needle approaches the eye, and coming closer together as it is moved farther away. The continuous lines represent rays from the needle, the interrupted lines rays from a nearer point in focus.

of transparent fibers from the anterior part of the hyaloid membrane 1.5 mm. or so in front of the ora serrata. Passing forwards they form a series of bundles which occupy grooves between the ciliary processes, and then bridge the gap, as just mentioned, between the ciliary processes and the lens. Near the lens circumference the zonula splits into an anterior and a posterior lamina, the former is the thicker of the two and blends with the lens capsule a little in front of the equator, and

constitutes what is generally known as the *suspensory ligament* of the lens. The space between the two layers of the zonula is called the *canal of Petit*. Slits in the layers of the membrane establish communications between the canal of Petit and the anterior chamber and the region behind the lens (post lenticular space, p 1163).

Light rays from an object at infinity, which is taken as any point more than 20 feet (6 meters) distant, are parallel and are brought to a focus (principal focus) on the retina of the emmetropic eye. Rays from a near object are divergent, but they too are brought to a sharp focus. This adjustment of the dioptrics of the eye whereby it is able to focus the image of both far and near objects is called *accommodation*. That the refracting power of the eye does actually undergo a change when it is turned from a far to a near object or vice versa was shown by Scheiner (1619) by a simple experiment. A card with two pin-holes separated by a distance less than the diameter of the pupil is held before one eye. The eye is focussed upon a needle held in front of the card and perpendicular to a line joining the two holes. The needle appears single, but it appears double if the eye is focussed upon an object placed either beyond it or between it and the eye. The explanation will be evident from fig 75.34.

There are at least three possible means by which accommodation of the eye could be brought about. The retina might be moved towards or away from the lens, i.e., the eye might be elongated or shortened so that divergent rays in the one instance or parallel rays in the other would be accurately focussed. This mechanism is actually made use of in the mollusc pecten. That it is not the method followed by the human eye was proved by Young (1801). A second possibility is that the distance between the retina and the lens is altered by a movement of the lens, this is the method used in photography, the distance between the film and the lens can be nicely adjusted for the focus of near objects. In the bony fishes accommodation is effected in such a manner⁴.

⁴ These fish are myopic when the eye is at rest, i.e., the eye is adjusted for near vision, accommodation for far vision is an active process consisting of contraction of a structure called the *campanula* which moves the lens backward. In some birds the central part of the anterior surface of the lens moves forwards into the pupillary aperture and as a result of pressure against the rigid margin of the iris becomes highly convex. In other avian species the cornea consists of two lamellae, the posterior being drawn backwards during accommo-

Variations in the convexity of the crystalline lens and consequently of its converging power is the third possibility (see fig 75 35) This is the method first suggested by Young for the human eye, the conception was later elaborated by Helmholtz That such is the mechanism adopted by mammals in general is now almost universally accepted A change in the convexity of the anterior surface of the lens during accommodation is a well-established fact and one which can be demon-

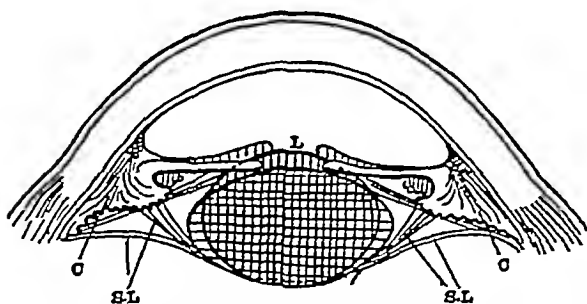


FIG 75 35 Illustrating the mechanism of accommodation of the eye for near vision The horizontally shaded lens and the unshaded iris show the position of the parts when at rest, the vertically shaded lens and iris show the position during accommodation for a near point. C, ciliary muscle, S.L, suspensory ligament. (Redrawn from Landolt.)

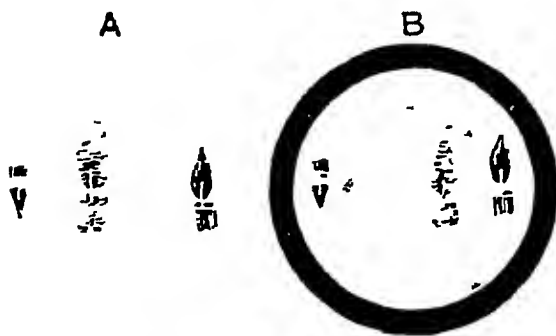


FIG 75.36 Purkinje-Sanson images A, during far vision, B, during accommodation for near vision (Redrawn and modified from Williams)

strated by the following experiments A lighted candle is held to the outer side and a little in front of the eye of a subject in a darkened room Three images (Purkinje-Sanson images, fig 75 36) within the subject's pupil will be seen by an observer, one bright and erect reflected from the cornea, another larger, erect and dimmer from the anterior surface (epithelial layer) of the lens which like the cornea acts as a convex mirror The third image is inverted, bright and smaller than the other two,

dation for far vision In others again (owls and hawks) the curvature of the cornea is increased during near vision

it is reflected from the posterior (concave) surface of the lens The subject is directed to gaze into the distance while the positions of the images are noted, he then looks at a near object when a change in the size and position of the reflection from the anterior surface of the lens will be observed It becomes smaller and moves towards the corneal image which, of course, remains stationary, as does also the inverted image from the posterior surface of the lens The change in size and position of the large erect image must mean that the anterior surface of the lens has become more convex ⁶ Now, if one knows the radius of curvature of the cornea, which can be measured by means of an instrument known as an ophthalmometer, then the radius of curvature of the surfaces of the lens and the changes in their curvature during accommodation can be calculated from careful comparative measurements of the sizes of the images from the cornea and lens The average values for the radius of the anterior surface of the lens in five subjects examined by Fincham were 12.2 mm for the "resting eye" and 6.8 mm during accommodation for near vision The change in curvature of the posterior lens surface was slight (about 0.5 mm) The average increase in thickness of the central part of the lens was 0.47 mm, while the equatorial diameter diminished by 0.5 mm The center of the anterior surface moved forward by from 0.3 to 0.4 mm In general terms the change in shape of the lens during accommodation for near vision can be summed up as follows The central part of the anterior surface becomes more convex, the posterior surface shows little change The peripheral region of the anterior surface actually becomes somewhat flatter, this surface taken as a whole therefore assumes a hyperbolic form

The manner in which the change in shape of the anterior surface of the lens is brought about was

⁶ The *phakoscope*, an instrument invented by Helmholtz, may be employed for observing the images It consists of a small hexagonal dark box (fig 75.37) with an aperture (A) for the examiner's eye and another for the eye of the subject. A candle placed in front of the prisms, P and P' throws a pair of images (squares of light) upon each of the three reflecting surfaces of the observed eye The aperture for the subject's eye is in the side of the box opposite to the window C While his eye is being observed the subject gazes into the distance and then accommodates for near vision by fixating a needle placed in front of the window This method of observation has the advantage that it is easier to detect a change in the distance between a pair of squares (which, of course, will indicate a corresponding change in size of the double image) than a change in the size and movement of a single image such as that of a candle flame

explained by Helmholtz as follows. When the eye is accommodated for distant vision the suspensory ligament which, as we have seen, is attached to the lens capsule, is drawn taut as a result of the pull of the elastic structures, e.g., the ciliary body and choroid. The peripherally directed traction exerted upon the lens capsule through the suspensory ligament results in flattening of the curvature of the anterior surface of the lens. Focussing the image of a near object is accomplished by contraction of the ciliary muscle which, by drawing the choroid forward, permits the ciliary processes to move forward and inward, thus reducing the diameter of the ring (corona ciliaris) which they form. The

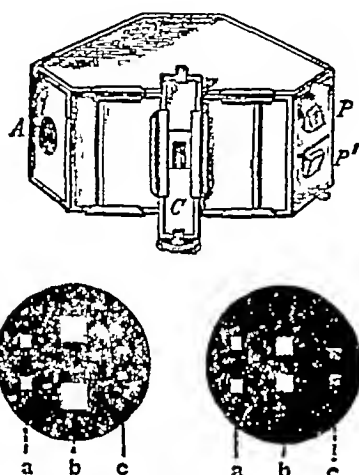


FIG 75.37 The phakoscope. Description in text (footnote 5) (After Helmholtz.) Reflections from cornea (a), anterior surface of the lens (b) and posterior surface of the lens (c) during far (left) and near vision (right).

suspensory ligament and lens capsule are thus relaxed, and the lens, by virtue of its inherent elasticity, assumes a more convex form (figs 75.35 and 75.38). The excised lens, i.e., one released from the restraint of surrounding structures, is therefore at its maximum convexity. Helmholtz's conception is supported in its main tenets by modern work. The movement inward of the ciliary processes during accommodation for near vision has been observed in the living human eye, and in an eye from which the lens substance had been adsorbed as a result of injury, tightening and slackening of the empty capsule was seen during the corresponding phases of accommodation. Yet the curvature assumed by the anterior surface of the lens is, as mentioned above, *hyperbolic*, whereas one would

expect it to assume a spheroid form were the lens substance itself elastic and the change in shape due simply to its recoil when released from restraint.

The details of the mechanism with respect to this point have been elucidated by the work of Fincham. According to this observer, the change in shape of the lens during accommodation is explained by the high degree of elasticity possessed by the lens capsule and by the regional variations in its thickness (p. 1153), together with the pliable nature of the cortical part of the lens as compared with the nucleus. When the lens is accommodated for distant vision, its substance is confined under tension within the capsule and, as a consequence, distends the latter to the greatest degree where it is weakest, (i.e., thinnest) namely, on the posterior surface. The convexity of this surface is therefore near its maximum when the eye is adjusted for

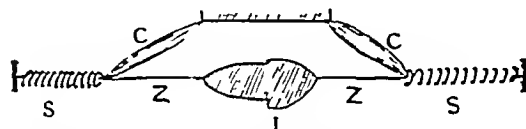


FIG 75.38 Diagram to illustrate the mechanism of accommodation. C, ciliary muscle, relaxed (left), contracted (right), S, spring representing the elastic choroid, L, lens, left section, for far vision, right for near vision, Z, zonula.

distant objects and little further change can occur during accommodation for near vision. Upon contraction of the ciliary muscle and the consequent slackening of the suspensory ligament the recoil of the elastic capsule moulds the plastic cortex, the peripheral part of the anterior surface is thus pressed back by the relatively thick capsule but the thinner central part of the latter permits the lens substance, chiefly the highly convex nucleus, to bulge forwards. That is, the anterior surface of the lens becomes somewhat "conoid" (Fincham).

THE VISUAL AXES AND PUPILLARY DIAMETER DURING ACCOMMODATION. The complete act of accommodation for a near object comprises, besides an increase in the convexity of the anterior surface of the lens, *convergence of the eyes* and *constriction of the pupil*. The constriction of the pupil which occurs during accommodation serves three purposes. The narrowed aperture of the iris (a) reduces chromatic and spherical aberration, thus increasing visual acuity, (b) diminishes the quantity of light (which is relatively greater from near objects) entering the eye and (c) increases the *depth of focus*.

The depth of focus of any lens system is defined as the greatest distance through which an object may be moved and still be sharply focussed. When a clear image is formed by a lens, each point in the object is, as a result of diffraction, a series of small concentric circles of light, rather than a geometrical point. When the distance from the lens is increased the image is formed in front of its previous position (that is, in front of the surface, such as a photographic film or retina), if the distance is reduced, the image is formed behind. In either case the rings of light upon the surface become larger. If they are of such a size as to still remain confined, each to a particle making up the grain of the surface, film or retina, the image is sharp and appears in true focus, if they are large enough to spread to neighboring particles the image of the object is blurred. In the case of the fovea the "grain" is determined by the cones. The depth of focus of the eye is, therefore, given as the greatest distance through which a point can be moved while its image remains restricted to a single cone. Thus, if a point which forms a sharp image is moved from a distance of 100 feet to one 125 or to one of 90 feet and remains just as clear but becomes blurred at greater or smaller distances, the depth of focus is 35 feet, all objects between the distances of 90 and 125 feet will be in focus. With the eye accommodated for near vision the depth of focus varies inversely with the diameter of the pupil. With a pupillary diameter of 4 mm and at a distance from the eye of 25 cm the depth of focus is less than 1 cm, whereas constriction of the pupil to a diameter of 1 mm increases it to about 3 cm. The depth of focus decreases as the distance from the eye is reduced. This tendency is, therefore, counteracted by the constriction of the pupil which occurs as part of the mechanism of accommodation of the eye for near vision.

The range and amplitude of accommodation The farthest point from the eye at which an object can be seen clearly is called the *far point* or *punctum remotum*. The corresponding point nearest the eye is termed the *near point* or *punctum proximum*. In the emmetropic eye the far point is at infinity, i.e., at a distance of over 20 feet (6 meters) and the near point at from 7 cm to 40 cm, depending upon age (fig 75.39). The difference between the far and the near point distances is termed the *range of accommodation*. The difference between the refracting power of the eye when accommodation is completely relaxed for the far point and fully displaced for the near point is called the *amplitude of accommodation*. The far point is con-

jugate to a point on the retina, i.e., parallel rays entering the eye come to a focus on the retina and rays from the latter upon emerging from the eye are parallel and would meet at infinity. Similarly, in the accommodated eye the near point is conjugate with a point on the retina. The focal length of the eye in each state of accommodation therefore corresponds, respectively, to the far and near point distances. It will be recalled that the refractive power is expressed as the reciprocal of the focal distance, the unit being 1 meter and called a diopter. The reciprocal of the far point distance is termed the *static refraction* (designated R) of the eye, and that of the near point distance, the *dynamic refraction* (P). The difference between the two (P-R) gives in diopters the amplitude of accommodation. In the emmetropic eye,

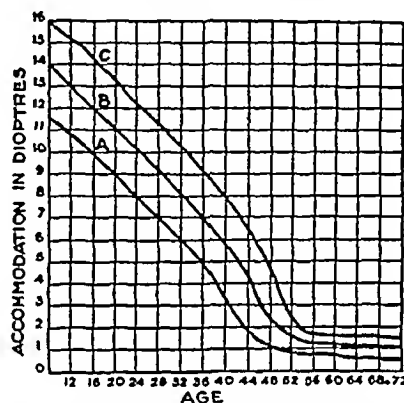


FIG 75.39 The amplitude of accommodation at different ages. A The lowest physiological values. B Average values. C Maximum values. (After Duane.)

since the far point is at infinity, the static refraction is taken as zero. When the near point is at 10 cm the dynamic refraction is $\frac{100 \text{ (1 meter)}}{10} = 10 \text{ D}$. The amplitude of accommodation in such an emmetropic eye is therefore 10 D.

The amplitude of accommodation diminishes progressively from childhood to about sixty years of age, being 16 D at twelve years, 6.5 D at the age of thirty and only about 1 D at sixty. In other words, with advancing years the near point gradually recedes from the eye and at sixty years of age an object must be 1 meter distant in order to be clearly focussed upon the retina (see *presbyopia*). This phenomenon is due in the majority of instances to physical changes in the properties of the lens and its capsule (reduced plasticity of the one or diminished elasticity of the other), it occasionally results from weakness of the ciliary muscle.

THE METABOLISM AND COMPOSITION OF THE LENS The lens is entirely epithelial in structure and devoid of blood vessels, it receives its nourishment from the fluids bathing its surfaces, e.g., the aqueous humor and the fluid of the vitreous. The capsule acts as a semi-

permeable membrane which separates substances of high osmotic pressure (equiv 1.2% NaCl) within the lens from the osmotically less active aqueous. During accommodation, when the pressure within the lens is around that of the general intraocular pressure, fluid presumably would as a result of this higher osmotic pressure pass inwards across the capsule, on the other hand, when the ciliary muscle is relaxed a reverse movement of fluid is to be expected, for the tension exerted upon the capsule will, by raising the intra-lenticular pressure reverse the pressure relationships and drive fluid out.

As one might expect the metabolism of the lens is low, Adams has demonstrated a small but definite oxygen uptake (30 cu mm per gram of lens per hour). An autoxidation system has been shown by Goldschmidt and by Adams, this consists of (1) *glutathione* which is in high concentration in the lens as compared with other tissues and (2) a *flavonoid protein residue* identified as β -crystalline. These two are maintained in chemical equilibrium by an oxidation reduction of the SH \rightarrow SS type. Though other tissues possess such an autoxidative system, in them it plays a secondary rôle, whereas in the respiration of the lens it is believed to be of paramount importance.

Four types of protein have been identified in the lens substance, (a) a *euglobin* or "albuminoid" (17%) which is water soluble, and two water soluble proteins—*pseudoglobulins*, (b) α -crystalline (11%) and (c) β -crystalline (6%), together with (d) a small quantity of albumin (0.2%). The albuminoid is present mainly in the nucleus, α -crystalline chiefly in the superficial part of the cortex and β -crystalline in the deeper parts. The two crystallines are particularly rich in the sulphur-containing amino-acids cystine and cysteine. The lens proteins as first shown by Uhlenhuth are *organ specific* not species specific, thus differing in their immunological behavior from red cells (p 63) and blood serum. For example, a solution of lens protein when injected into an animal of the same or of another species causes the production of an antibody—a *precipitin*. This anti-serum has then the power to precipitate a solution of lens protein from wherever derived, i.e., from a species other than the one which supplied the anti-serum, from the same species or even from the same animal.

The lens substance contains a high concentration of potassium—400 mg per 100 grams of wet weight—as compared with about 3 mg per cent in the aqueous and 20 mg per cent in serum. The concentrations of calcium (5 mg per cent), sodium chloride (300 mg per cent) magnesium and silicates are relatively low. The total salt concentration is between 0.7 and 0.8 per cent. Cholesterol and phosphatides amount to about 200 mg per cent in young lenses, but are from two to four times this value in older specimens.

CHANGES IN THE LENS WITH AGE—CATARACT
The loss of plasticity and elasticity of the lens

with age as a result of a gradual sclerosis, and the effect such changes have on the mechanism of accommodation are referred to on p 1116. Some alterations in lenticular color may accompany the sclerosing process, amber tinting, or even a reddish or brownish discoloration of the lens with consequent filtering of the shorter rays, is of common occurrence.

Cataract is the name given to any partial or complete opacity of the lens. In the commonest variety no ocular or general disease which can be held responsible precedes the development of the opacity, and thus, since it appears to be simply a manifestation of age, is termed *senile cataract*. The process leading to the opacity is degenerative in nature, not inflammatory, for the lens is, as just stated, avascular. The opacity commences usually in the deeper part of the cortex and does not, as a rule, involve the nucleus. The lens swells as a result of accumulation of fluid between the fibers, the anterior chamber becoming shallow. So long as the superficial layers of the cortex are clear the cataract is called *immature*. It is said to be *mature* when the opacity has extended to include the superficial layers.⁴ The water content of the lens has by this time returned to normal. The mature stage is followed by disintegration of the cortex which becomes softened into a pulsatious mass, this is the stage of hypermaturity, drying and shrinkage of the lens finally result.

The essential change in the cataractous lens is a progressive coagulation of the lens proteins. According to the most generally accepted explanation, such a process is due to the prolonged action of ultraviolet light, and, in some instances, to the thermal effect of infra red rays. The lens as already pointed out, by absorbing a large proportion of the rays below λ 350 m μ and 400 m μ , protects the retina from their injurious effects. Wave lengths below 295 m μ are absorbed by the cornea. The rays absorbed by the lens are not without their effect upon the lens substance itself, it is these which cause the physical change in the lens proteins. Two stages are recognized in the coagulation process: (1) *denaturation of the lens proteins*, consisting presumably of a molecular rearrangement, by light or heat which renders them susceptible to (2) *aggregation (agglutination) of the protein particles* into a flocculent mass—coagulation. This ultimate result occurs only in the presence of certain salts and is enhanced by some organic substances, e.g. dextrose and acetone. The theory that radiant energy is responsible for denaturation of the lens proteins is in accord with many observations. For example, the absorption by the lens

⁴ As long as a clear interval exists between the opacity and the iris (which lies in contact with the anterior surface of the lens) the latter throws a shadow upon the former. When the opacity involves the entire thickness of the cortex, i.e., right up to the iris, there is no shadow.

of the shorter rays increases with age. The opacity commences in the lower quadrant of the lens which receives the most intense light. In tropical countries, e.g. India and Egypt, cataract is much commoner than in temperate latitudes, it is also less frequent in the latter than in Arctic zones, presumably as a result of the high content in actinic rays of the light reflected from snow and ice. It is also stated that on this continent the incidence of cataract increases from temperate zones to the equator, and that it is also higher in those who work in the fields than in city dwellers. Burge's experiments and the more recent ones of Clarke show convincingly the effect of light upon the development of lenticular opacities. Burge found that whereas exposure of a solution of lens protein to ultraviolet light for 100 hrs did not cause coagulation, this occurred if CaCl_2 , MgCl_2 , dextrose or acetone were added. Moreover the exposure of the eye of a living fish or frog to short light waves was without effect if the animal had previously been kept in tap water, but definite opacity of the lens followed a few hours exposure, if the fish or frog had been for some days in water containing 0.8% calcium chloride, 0.1% dextrose or 0.1% sodium silicate. Clarke found that heat enhanced the action of the light rays upon solutions of lens proteins, and that opacity could not be produced in the absence of calcium.

The incidence of cataract is much higher in diabetics than in normal persons, it is usually of the ordinary or so-called senile type, but it occurs at an earlier age. The opacity is attributed to the action of dextrose and possibly of acetone bodies in rendering the lens proteins more readily coagulable by light. Duke Elder believes that as a result of the high blood sugar, the osmotic relationships between the lens and the surrounding fluids are disturbed and the nutrition of the lens thereby interfered with.

Though the factors outlined above appear to be the main ones concerned in the production of the common or senile type of cataract, opinions differ considerably as to the details of the mechanism involved, as well as in regard to the production of other types of lenticular opacity. It is suggested, for example, that the action of ultraviolet rays in inducing denaturation of the lens proteins is due to the reduction of the glutathione and β -crystalline content and the consequent depression of the autoxidative mechanism. Such an effect of ultraviolet radiation has been shown experimentally, it has also been established that the content of the lens in glutathione and in the thermostable protein residue diminishes with age. The power of the lens to fluoresce upon exposure to short wave radiations (p. 1116) and the disposal of the surplus energy by converting them into long waves, is considered by Burge to be an important factor in ameliorating the effect of light upon the lens, fluorescing bacteria for example are much less readily killed by ultraviolet light than are other types. It is of some considerable interest therefore that the power of the lens to fluoresce diminishes with age.

In glass blowers' cataract, infra-red rays would appear to play an important auxiliary rôle. Hartridge and Hill believe that these act indirectly by increasing the production of aqueous humor by the ciliary processes and iris which leads to disturbances in the fluid interchanges between the lens and its surroundings, the nutrition of the lens suffering as a result. They point out that the lens absorbs only about 12% of the heat rays which enter the eye whereas the iris absorbs all which reach it, it is unlikely, therefore, that their action upon the lens proteins is direct. Heat rays may act, however, simply by accelerating the denaturing action of ultraviolet light (Clarke). Duke Elder, on the other hand, suggests that the heat rays affect the permeability of the capsule, with consequent upset in the osmotic relationship between the lens substance and the aqueous.

An interesting type of cataract is that following parathyroidectomy, disordered calcium metabolism would appear to be in some way concerned in its production, the calcium content of the lens is increased, whereas that of the blood is reduced. There is no definite evidence, however, to connect ordinary senile cataract with parathyroid deficiency, but there seems little doubt that calcium is concerned in some obscure way with cataract development. A favorite method of producing cataract for experimental study is by the injection of naphthaline, opacities form only if the animal is on a low calcium diet. In rats, lenticular opacity is readily produced by a diet containing a high percentage of lactose or galactose, the former sugar increases the absorption of calcium, but whether this fact has any bearing upon the development of the cataract is difficult to say, it does not dovetail with other observations in respect to calcium and the development of cataract.

The composition of the cataractous lens shows marked differences from the normal. There is an increase in the insoluble albuminoid and a decrease of the soluble proteins and of glutathione, the oxygen uptake is much reduced. Of the inorganic constituents calcium shows a relatively enormous increase (up to 140 mg per cent). The concentration of magnesium, sodium and silicates is also raised, whereas that of potassium is greatly reduced.

OPTICAL DEFECTS

SPHERICAL ABERRATION Rays traversing the peripheral parts of an ordinary convex lens are refracted more strongly and therefore come to a focus nearer the lens than do those transmitted through more central regions. In other words, the outer and inner rays cross in front of the retina and a blurred image is formed (fig. 75.40).⁷ This is an

⁷This is called spherical aberration "with the rule". But sometimes, though much less commonly, the central rays are refracted the more strongly, this is termed "against the rule". The spherical aberration of the

inherent defect of convex lenses and is called *spherical aberration*, in the manufacture of a camera lens, special means are employed to correct it, the lens being built up of separate pieces of glass of different refractive indices cemented together so that all rays are converged to the same point. Spherical aberration is corrected to a certain extent and in a somewhat similar manner in the crystalline lens, the nucleus having a higher refractive power than the periphery (see p. 1145). The iris, since it covers the outer part of the lens and shuts off the peripheral rays, also serves to correct this defect.

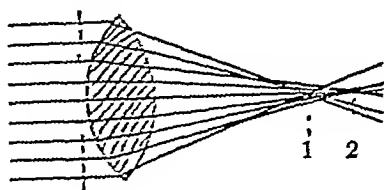


FIG. 75-40 Spherical aberration. Outer rays meet at 1, inner rays at 2.

CHROMATIC ABERRATION. The colors composing white light are refracted to different degrees according to their wave lengths. The violet rays are refracted most, the refractibility diminishing progressively from the violet to the red end of the spectrum (fig. 75-41). For this reason a series of fringes, colored from violet to red from within outwards, borders the image formed by a simple convex lens. *Chromatic aberration*, as this defect is called, is corrected in camera and microscope lenses by cementing a biconvex lens of crown glass to a concave one of flint glass. Such a lens is called *achromatic*. The eye is only partially corrected for chromatic aberration. The yellow rays are focused upon the retina, but the short blue rays being more acutely bent meet in front, while the longer less strongly refracted red rays come to a focus behind the retina. Other rays are bent similarly in accordance with their wavelengths. Furthermore, the retina being on the temporal side of the optical axis, the blue rays fall upon the retina nearer to the point (*axial point*) where it is cut by the optical axis. Thus, of course, will obtain in both eyes, and lines projected outwards from the axial points will meet as shown in figure 75-41. The blue rays will

therefore appear to come from a point farther away than the point emitting the red rays; the image formed on the retina by the shorter blue rays will also be smaller than that formed by the red rays. These effects of the unequal refraction of the different colors are called respectively *anomaloscopic* and *chromatic difference of magnification*. In ordinary vision the colors surrounding the images on the retina are not perceived, we have come to ignore them.

Diffraction of light or scattering of light. Light is diffracted by the pupillary margin and by the lens fibers and cortical epithelium, as a result of this and of spherical aberration the retinal image is not made up of points of light but of diffusion circles (blur circles), i.e., a bright central disc surrounded by light rings which diminish in intensity by almost imperceptible gradations toward the periphery. The relative size of the

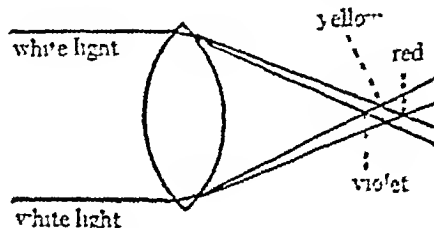


FIG. 75-41 Chromatic aberration

central bright area varies inversely with the diameter of the pupil and directly with the wave length of the light. The opposite effects of changes in the size of the pupil upon this defect and upon chromatic aberration have been pointed out (p. 1155).

None of the ocular media is perfectly homogeneous, owing to their colloidal nature a certain proportion of the light entering the eye is scattered (Tyndall phenomenon), that is, it is not focussed upon the retina but is deflected from the course which it would follow according to the laws of refraction if the contents of the globe were perfectly transparent. The colloidal particles have a size of the order of the wave length of light. The quantity of scattered light is directly proportional to the square of the size of the particles and inversely proportional to the fourth power of the wave length ($R \propto 1/\lambda^4$). Thus the greatest scattering within the eyeball occurs with violet and ultra-violet light and the least with red. The dispersion of ultraviolet rays probably serves a useful purpose in that the retina is thus protected from their injurious effects.

PRESEBYOPIA (*Gr. presbys*, old, *ops*, the eye) is the term given to the gradual reduction in the amplitude of accommodation on which goes hand in hand with advancing years.

EMMETROPIA AND AMETROPIA. The four optical

cornea is "with the rule" whereas that of the crystalline lens is "against the rule", the two types of spherical aberration, therefore tend to neutralize one another, and the spherical aberration of the eye, as a whole, will be the resultant of the two.

defects just described may be regarded as physiological, the first three being inherent to some extent in optical systems in general, the fourth is a natural accompaniment of age. Two other defects of frequent occurrence are due to incongruity between the length of the eyeball and its refracting power and must be classed as definite abnormalities.

The refractive state of the normal eye, which has its far point (p 1157) at infinity, i.e., at a distance greater than 6 meters (20 feet), is called *emmetropia*. Parallel rays entering the emmetropic eye are brought to a clear focus on the retina without any effort of accommodation. The static refraction (p 1157) of such an eye is therefore zero. If the far point is not at infinity the eye is *ametropic*. There are two forms of ametropia—myopia and hypermetropia (fig 75 42)

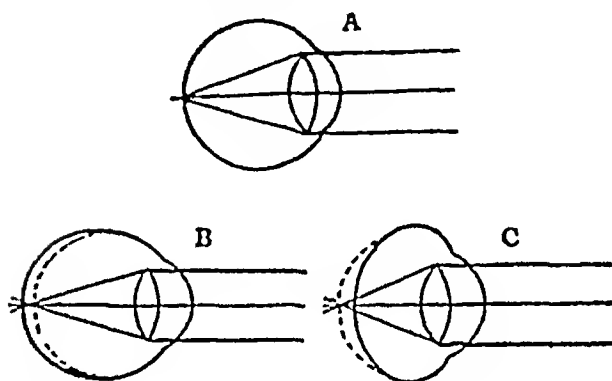


FIG 75 42 A, emmetropia, B, myopia, C, hypermetropia

In *myopia* (Gr *nyo*, I blink or half close the eye, *ops*, the eye) or *short sight*, the eyeball is too long relatively to its refracting power (fig 75 42, B). Obviously, such an eye will bring parallel rays to a focus in front of the retina, i.e., in the vitreous. After meeting, the rays cross and form a blurred image or a diffusion circle upon the retina, just as a camera which is extended too far forms an indistinct image upon the film. In order to form a clear image on the retina of the myopic eye the rays must be, not parallel but divergent, such as are emitted by a near object. The far point is therefore at a finite distance, and in extreme instances may be only a few centimeters from the eye. Accommodation, of course, is relaxed for the far point as in the emmetropic eye for, obviously, increasing the converging power of the lens will only cause greater blurring of the image. The far and near points being close together, the range and the amplitude of accommodation (p 1157) are reduced. Myopia is corrected by means of concave (diverg-

ing) lenses. If, for example, an object can be seen clearly no farther away than 1 meter, the myopia is -1 D, i.e., a concave lens of this power is required.

In *hypermetropia* (Gr *hyper*, above, *metros*, measure, *ops*, the eye) or *long sight* (fig 75 42C) the eye is too short for its refracting power.⁸ Therefore parallel rays after refraction fall upon the retina before they have come to a focus, and form an image blurred by diffusion circles. The rays as they emerge from the eye are divergent. The far point is *virtual*, i.e., it is behind the eye at the point where the rays would meet if continued backwards through the retina. The hypermetrope must accommodate when he views distant objects in order to focus the parallel rays upon the retina. The range of his accommodation (p 1157) is the

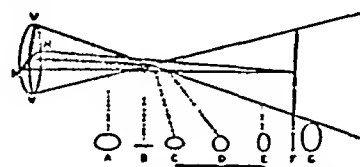


FIG 75 43 Refraction by an astigmatic lens. VV, the vertical meridian of the refracting body, is more curved than HH, the horizontal meridian. A, B, C, D, E, F, G show different sections of the beam after refraction. At B the vertical rays are brought to a focus, at F the horizontal rays are brought to a focus. From B to F is the focal interval.

same as that of the emmetrope but the amplitude is greater. Suppose, for example, that the near point is 0.10 meter in front of the eye and the far point is 0.25 meter behind, i.e., negative (-0.25). Then the dynamic refraction is $100/10 = 10$ D, and the static refraction is $-100/-25 = -4$ D, the amplitude of accommodation is therefore 10 D $- (-4$ D), or 10 D $+ 4$ D $= 14$ D. Hypermetropia is corrected by means of a convex lens, the distance of the far point behind the eye giving the measure of the strength of lens required, if this is -0.25 meter then a $+4$ D lens would correct the defect (see p 1149). The larger amplitude of accommodation in hypermetropia is accompanied by hypertrophy of the ciliary muscle, in myopia, on the other hand, the muscle shows atrophy and the circular fibers may be absent.

ASTIGMATISM (Gr *a*, privative, *stigma*, a point). In this condition, as its name implies, rays of light are not brought to sharp points upon the retina,

⁸ Both types of ametropia in the great majority of instances are due to an abnormality in the length of the eyeball and not to any change in the refracting power of the eye.

but form short lines. The defect is present in all eyes to a certain degree, when moderate it is therefore physiological. Only when the fault seriously reduces visual acuity is it abnormal. It must be remembered that rays of light pass through all meridians of a lens, in converging to a focus they therefore form a cone of light, not simply a flat pennant like beam. If all meridians of a lens have the same curvature, then rays in all planes will be refracted to the same degree and come to a focus together. If, on the other hand, the curvatures differ, the rays transmitted through a meridian with the greater curvature will be refracted more strongly and brought to a focus in front of rays passing through other meridians. For example,

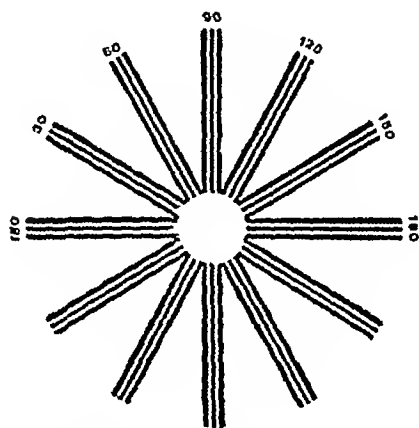


FIG 75 44 Chart used in testing for astigmatism. One eye being closed the patient is asked to say which of the groups of lines are blackest and most distinct and which are lightest and indistinct.

should the vertical meridian be more curved than the horizontal, then when the rays passing through the vertical meridian are in focus those in the horizontal will form, not a point, but an ellipse, a circle or a line. Thus there are two foci and the distance between two lines—the *focal interval*—is a measure of the degree of the astigmatism (see fig 75 43). Such inequalities of curvature in the meridians of the cornea or, less commonly, of the crystalline lens, are the cause of what is known as *curvature astigmatism*. The greater curvature may be in either the vertical, horizontal or an oblique meridian. But in physiological astigmatism (which is due most probably to the pressure of the upper lid upon the eyeball during growth) the greater curvature is in the vertical meridian. In pathological types also this meridian has most commonly the greater curvature. When the refractive power

though unequal in the various meridians is the same throughout any meridian, and the maximally and minimally curved meridians (*principal meridians*) are at right angles to one another, the astigmatism is called *regular*. In *irregular astigmatism* not only is the refractive power unequal in the different meridians, but it is not uniform throughout a meridian, and the principal meridians are not at right angles to one another. Astigmatism may result from inequalities in the refractive indices between different parts of the lens, this is termed *index astigmatism*. Slight inequalities of this nature in the refractive indices may also contribute to physiological astigmatism. An oblique or other malposition of the lens is sometimes a cause of astigmatism. When the subject of astigmatism looks at a clock face, the straight lines in the vertical numerals XII and VI may be clearly seen, while the horizontal lines in IX and III are blurred, or vice versa. Or, the diagonal numerals may be out of focus while the vertical and horizontal are sharply defined. A chart such as is shown in figure 75 44 is employed to detect the meridian or meridians in which the corneal curvature is abnormal.

Astigmatism is corrected by the use of spectacle lenses convex in the meridian corresponding to that of the cornea (or crystalline lens) having the lesser curvature. Thus if the curvature of the cornea is greater in the vertical meridian, the subject is fitted with a cylindrical lens, which, it will be recalled, refracts in a single plane, having its convexity in the horizontal meridian.

Contact lenses

The contact lens consists of a thin cup-like shell of glass moulded to fit the cornea and sclera and ground to the required curvature. It is applied directly to the eye and therefore moves with it. Contact lenses are employed for the correction of various types of refractive error, e.g., myopia and hypermetropia and those due to abnormalities in the form of the cornea (astigmatism or conical cornea) or to absence of the lens (aphakia). Before being applied the concavity of the lens is filled with saline solution and when in position is separated from the cornea only by a thin film of the solution. The refractive indices of the cornea and the saline are approximately the same. The two therefore constitute a single refracting medium moulded as it were by the posterior surface of the glass shell into a normal form. Light rays are refracted at the anterior surface of the cornea in contact with the air and less strongly at the posterior surface, but undergo no further refraction until they reach the crystalline lens (fig 75 45).

THE INTRA-OCULAR FLUIDS

THE INTRA-OCULAR PRESSURE The pressure within the chambers of the eye of a living animal is from 20 to 25 mm Hg and from 8 mm to 10 mm Hg in the excised but intact globe, it also falls to the latter level immediately after death or after arrest of the ocular circulation. The difference between these two sets of values, is due to the pressure of the blood in the vessels of the globe. This will be clear from fig 75 46. The vascular bed of the eyeball lies between the relatively resistant sclerotic on the outside and the incompressible intra-ocular contents on the other. It is evident that blood pumped into the vessels of the choroid will raise the intra-ocular pressure above that of a bloodless eye. The intra-ocular pressure runs closely parallel with that of the blood in the choroidal capillaries. A rise or fall in general arterial pressure therefore may cause a corresponding change in intra-ocular pressure, though it will be much less in degree. The pulse beat causes a variation in intra-ocular pressure of from 1 to 2 mm Hg, and the respiration one of from 3 to 5 mm Hg. However, since it is the pressure in the capillary bed of the eye rather than that in the larger ocular vessels which is the determining factor and the former pressure can vary as a result of *local* changes in caliber of the minute vessels, the intra-ocular and arterial pressures do not necessarily change in the same direction. Thus arterial hypertension, in which the capillary pressure is not raised (p 163), does not cause a rise in intra-ocular pressure, amyl nitrite, on the other hand, causes a *fall* in arterial pressure as a result of the peripheral vasodilatation, accompanied by a *rise* in intra-ocular pressure (due to the dilatation of the capillaries of the eye and the consequent increase in capillary pressure). The intra-ocular pressure therefore follows very closely the venous pressure, tying the vortex veins, for example, increases capillary pressure and causes a rise in intra-ocular pressure of from 50 to 60 mm Hg. The pressure of 8 or 10 mm Hg which, as mentioned above, exists within an eye immediately after its circulation has been arrested, may be looked upon as representing the balance struck between the production of intra-ocular fluid and its removal through the drainage channels.

Other factors which affect the intra-ocular tension are, pressure from without by the action of the eyelids and of the extrinsic muscles of the eye. The effect exerted by the eye muscles in ordinary movements is

negligible, but maximally strong convergence of the eyeballs may raise the pressure by from 4 to 10 mm Hg. Movements of the lids have a greater influence, strong contraction of the orbicularis oculi causing a rise of 50 mm Hg or more. Exposure of the eye to light causes a fall, and dark adaptation a rise in pressure which is attributed to constriction and dilatation, respectively, of the ocular capillaries. Contraction of the ciliary muscle causes no change in the intra-ocular pressure, nor does the state of the pupil. A change in pressure does not occur therefore upon accommodation of the eye, provided that the associated convergent movement is prevented.



FIG 75 45 Contact lens

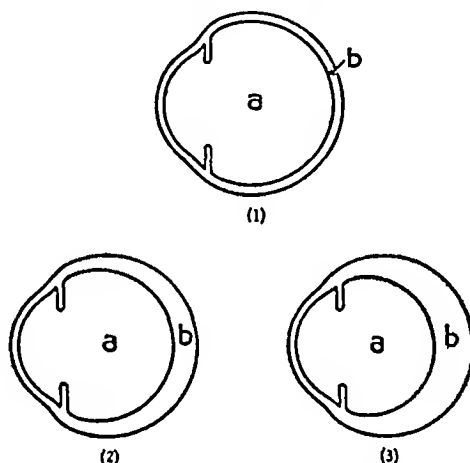


FIG 75 46 Showing the distribution of volume pressure in the eye with (1) a collapsed blood-pressure, (2) in the normal state, and (3) with raised blood pressure and capillary dilatation (After Duke-Elder The nature of the intra-ocular fluids)

THE VITREOUS BODY The vitreous body occupies the segment of the globe lying behind the lens and the ciliary processes. In front it presents a saucer-shaped depression—the *hyaloid fossa*—which lodges the posterior convexity of the lens. A narrow space (*post-lenticular space*) filled with fluid separates the lens from the concave surface of the vitreous. The *hyaloid canal* (which lodged the hyaloid artery in the embryo) runs from the posterior pole of the lens to the center of the optic disc. The structureless *hyaloid membrane* surrounds the vitreous, anteriorly it is strengthened by radial fibers which have already been referred to as the *zonula ciliaris* (p 1154). In this region it shows a series of grooves which lodge the ciliary processes. The vitreous body is a jelly-like material, possessing no obvious structure, but it is claimed that by the use of special fixatives (e.g., a weak solution of chromic acid)

a series of superimposed lamellae arranged concentrically around the hyaloid canal and composed of very thin flat cells can be demonstrated. The spaces between the lamellae contain fluid almost identical in composition with the aqueous humor. Though some maintain that the lamellae are fixation artefacts and that the vitreous body is a homogeneous gel, there seems little doubt that a framework does exist. Friedenwald and Stuehler observed it in the vitreous of ox eyes which had been prepared without the use of fixatives. The structural framework contains a protein called *vitron*. The refractive index of the vitreous is 1.33 (see p 1144).

THE AQUEOUS HUMOR is a clear watery fluid occupying the anterior and posterior chambers of the eye. It has a refractive index of 1.33, it is alkaline in reaction (pH 7.1-7.3), has a specific gravity of from 1002 to 1004 and a viscosity of 1.029. Its composition and that of serum are given in the following table, after Duke-Elder

	Grams per 100 cc Aqueous Humor	Serum
Water	99.6921	93.3238
Solids	1.0869	9.5362
Total protein	0.0201	7.3692
Albumen	0.0123	4.4135
Globulin	0.0078	2.955
Fats	0.004	0.13
Glucose	0.082	0.100
Urea	0.058	0.070
Sodium	0.365	0.378
Chloride	0.458	0.436
Potassium	0.023	0.025
Calcium	0.007	0.009
Magnesium	0.017	0.018

The chief differences between the compositions of serum and aqueous are shown in the concentrations of substances of colloidal nature—the proteins and fats. Glucose and urea are also in lower concentration in the aqueous humor than in the plasma, the osmotic pressure of the aqueous is lower as well.

THE FORMATION OF THE INTRA-OCULAR FLUIDS

Owing to the importance of the subject to glaucoma, the formation of the intra-ocular fluids has been investigated most diligently, especially by Duke-Elder and his associates. But the precise mechanism involved has not yet been established. There have been three theories of the process proposed, namely, that it is one of (1) *ultrafiltration*, (2) *dialysis* or (3) *secretion*. None of these theories alone fits all the known facts. The principal process appears to be one of ultrafiltration, and involves those factors,—capillary blood pressure and colloid osmotic pressures—which have been described elsewhere in connection with the interchange of water and substances in solution across capillary

membranes (chaps 3 and 13). The high capillary pressure required for ultrafiltration into the interior of the eyeball has been shown to exist (owing to the relatively high hydrostatic pressure therein), the mean pressure in the capillaries derived from the retinal artery and the posterior ciliary arteries is 50 mm. Hg which is much higher than that in the capillaries of most other regions. Nevertheless, ultrafiltration is not a complete explanation of the formation of intra-ocular fluid, for when the concentrations of the constituents of plasma and of the aqueous humor are compared, it is evident that the latter is not simply an ultrafiltrate like the fluid in Bowman's capsule, for example. It is therefore thought that dialysis plays a part, though there are discrepancies here which do not permit the formulation of a theory based upon simple dialysis alone. There is also some evidence that certain substances, e.g., sodium are secreted.

The surface of the ciliary body and the posterior aspect of the iris are believed to constitute the membrane interposed between the aqueous humor and the blood. That diffusion occurs across both these surfaces is indicated by the following observations. When the pupillary aperture is occluded either experimentally or as a result of disease, aqueous fluid accumulates behind the iris. On the other hand, fluid is formed after excision of the iris or when it is congenitally absent. In certain fish not possessing a ciliary body and in the congenital absence of the latter in man, normal aqueous humor is present, it has also been found in a cyst of the iris itself.

THE CIRCULATION OF THE INTRA-OCULAR FLUIDS
THE DRAINAGE SYSTEM OF THE EYE. Intra-ocular fluid is probably reabsorbed to some extent from all parts of the interior of the globe. A small proportion of the dialyzed fluid passes from the posterior chamber through the zonule and down the hyaloid canal to the lymphatics of the optic nerve. However, less than 1 per cent is reabsorbed in this way, or indeed from any region of the eye lying posterior to the iris. The chief exits for the fluid are at the angle of the anterior chamber (*angle of the iris* or *filtration angle*) and from the anterior surface of the iris. It will perhaps be of advantage to the reader if the main structural features of this region are recalled.

On the deep aspect of the sclera at its junction with the cornea and in front of the angle of the anterior chamber lies an annular venous sinus—the *canal of Schlemm* (*sinus venosus sclerae*) (fig 75.30). The sinus

completely surrounds the corneal margin and in meridional sections of the eye appears as a small oval gap or cleft lined by endothelium. The inner or posterior wall of the canal is separated from the anterior chamber by a zone of trabecular tissue formed by the breaking up of the posterior elastic lamina of the cornea and termed the *pectinate ligament*, the intervals between the trabeculae are termed the *spaces of Fontana*. The trabecular tissue of the pectinate ligament is continued around the iridial angle, its fibers terminating in the tissue of the iris. Schlemm's canal is fed by an *afferent arteriole* derived from the ciliary arteries and drained by an *efferent venule* which empties into the episcleral venous plexus (Friedenwald). The spaces of Fontana communicate with the anterior chamber, but there is no *direct* communication between the former and the lumen of Schlemm's canal, the canal cannot be injected, for example, with a colloidal solution introduced into the anterior chamber.

The *scleral spur* is the term applied to a small triangular projection of the sclera on the posterior aspect of the sclerocorneal junction, it lies immediately behind the outer part of the posterior wall of Schlemm's canal and gives attachment to the meridional fibers of the ciliary muscle. Contraction of the muscle, by pulling upon the spur, is said to dilate the canal and thus favor the drainage of fluid from the anterior chamber.

Movement of the intra-ocular fluid from the posterior to the anterior chamber and from the latter to the filtration angle is brought about largely through intermittent variations in intra-ocular pressure occasioned by the several factors already discussed (e.g., pulsatile and respiratory variations in blood pressure and actions of the eyelids, etc.). Temperature differences between the superficial and deeper parts of the anterior chamber (*thermal factor*) cause convection currents to be set up which also play an important part in the movement of fluid. At the angle of the iris the fluid percolates into the spaces of Fontana whence it is absorbed across the posterior wall of the canal of Schlemm.

Friedenwald and Pierce have demonstrated a differential absorption between the water, crystalloids and protein constituents of the aqueous. From their experiments which involved the introduction of substances into the anterior chamber, these observers conclude that crystalloids and a small quantity of water are reabsorbed from the anterior surface of the iris, passing by diffusion through the walls of the capillaries. Colloids are removed by the phagocytic action of the surface layer of epithelial cells. A part of the protein is hydrolyzed by the action of enzymes present in the

fluid and reabsorbed as amino-acids. Water is absorbed chiefly through the spaces of Fontana and the canal of Schlemm. The rate of passage of fluid through the wall of the canal is governed apparently by hydrostatic and osmotic forces. The pressure of blood in Schlemm's canal (or rather in the small veins leading from it) is stated to be equal to or about 1 mm higher than that of the fluid in the anterior chamber, but after the absorption of crystalloids one would expect the osmotic pressure of the aqueous to be considerably lower than that of the serum, under such circumstances an uptake of fluid would occur, provided the blood flow through the canal did not fall below a certain level. Slowing of the circulation would tend (as a result of the dilution of serum colloid dilution by the reabsorbed fluid and consequent reduction in osmotic pressure) to reduce the rate of reabsorption, an increase in blood velocity to increase it. Such a relationship was actually observed by Friedenwald and Pierce. Increase in intra-ocular pressure would quite evidently increase reabsorption, the pressure would thus tend automatically to be restored to its original level.⁹

GLAUCOMA OR OCULAR HYPERTENSION Persistent elevation of the intra-ocular pressure occurs as an accompaniment of several diseased states of the eye and may then be due to blockage of the drainage channels at the iridial angle or to the excessive production of fluid. The latter effect may result from mechanical irritation of the ciliary processes (e.g., by displacement of the lens) or obstruction of the venous channels with consequent rise in capillary pressure. Ocular hypertension associated with some such obvious disease of the eye is referred to as *secondary glaucoma*. When the intra-ocular pressure is persistently elevated above 35 mm or so and no abnormality of the eye exists to account for the hypertension, it is termed *primary glaucoma*. At the outset it may be said that, though primary glaucoma has been the subject of much speculation, its cause remains obscure.

The excessively high intra-ocular tension causes com-

⁹ According to some authorities (Maggiore, Duke-Elder) the canal of Schlemm contains blood only when the ocular venous pressure is inordinately high, being filled under usual circumstances with an aqueous fluid (with an osmotic pressure around that of the fluid in the anterior chamber). Since, as mentioned above, the pressure of blood in the small veins leading from it is higher than the normal intra-ocular pressure the canal could not serve as a pathway for the continued reabsorption of fluid, it is claimed that reabsorption can occur only if the intra-ocular pressure rises above the normal level. This conception attributes a safety-valve function to the canal, i.e., it is called into play presumably only in an emergency.

pression of the vessels and in time serious disturbances in the nutrition of the eye result, namely, optic atrophy, excavation ("cupping") of the disc, blindness and ultimately disintegration of the optical mechanism. Owing to the readjustments which take place in the ocular circulation the hypertension may exist for some time before any of these effects make their appearance. The pressure as it gradually rises first compresses the venous channels but, as a result of the opening up of the arterioles and capillaries, a larger proportion of the arterial pressure is transmitted to the venous side, the compressing force is thus overcome and the circulation maintained. This stage in the progress of the condition is referred to as *compensated glaucoma*. A point will be reached, however, at which the pressure of the intra-ocular fluids approaches equality with that in the ophthalmic artery, then further compensation becomes impossible and the structural changes just mentioned supervene. The condition is then termed *inflammatory*, or *decompensated glaucoma*.

The possible factors which have been suggested in explanation of the elevated pressure will be briefly considered. Mechanical obstruction at the filtration angle due to reduction in the depth of the anterior chamber and the consequent adhesion of the periphery of the iris to the cornea, is frequently present in decompensated cases, but this is secondary and not primarily related to the hypertension. It is not improbable that some abnormality in the nervous control of the vascular bed of the globe is fundamentally responsible. Friedenwald suggests that the reabsorption of water through the mechanism of Schlemm's canal as a result of sclerosis and narrowing of the afferent vessels (p. 1164) may be the essential factor concerned. Reduction in the

caliber of these vessels would tend, by slowing the blood flow through the canal, to reduce the reabsorption rate. Spasm of the vessels feeding the sinus would have a similar effect. On the other hand, the reduced depth of the anterior chamber in decompensated glaucoma suggests that the increased pressure originates in the posterior chamber of the eye and the theory has been advanced that swelling of the vitreous is responsible. From the gel-like nature of the vitreous body one might expect its water content to vary with changes in its inorganic constituents or in pH. It has been found, however, that the chemical changes necessary to cause any significant increase in volume of the vitreous are far greater than any that could occur in the body. Finally, a vasodilator toxin of the histamine type has been suggested which supposedly, by causing dilatation of the intra-ocular capillaries and an increase in the permeability of their walls, would lead to overproduction of intra-ocular fluid. The aqueous in decompensated glaucoma has a higher protein, and lower chloride content and osmotic pressure than normal, this fact lends some force to the theory of increased capillary permeability. The abnormal permeability would also, by reducing the effective osmotic difference between the contents of the anterior chamber and the blood, tend to diminish reabsorption. Nevertheless, analysis of the aqueous humor in *compensated* cases shows no significant departure from the normal and therefore lends no support to the idea that a change in capillary permeability is the fundamental factor in the development of glaucoma.

There is no causative relationship between arterial and ocular hypertension.

THE VISUAL FIELDS AND PATHWAY MOVEMENTS OF THE OCULAR MUSCLE STEREOSCOPIC VISION

THE VISUAL FIELDS The visual field of one eye is the part of the external world which is seen by that eye at any given moment, i.e., when its gaze is fixed in one direction. It may be likened to a portion of a great hollow sphere—a bowl—upon the interior surface of which the images of the external world are projected. Traquair pictures the visual field as “an island of vision surrounded by a sea of blindness”, carrying the simile further, the surface contour of this imaginary island is described in terms of visual acuity, the highest point corresponding to the fovea, a deep (bottomless) pit to the blind spot and a gradual slope toward the peripheral retina. The visual field of each eye subtends an angle of about 160° in the horizontal and 145° in the vertical meridian. The visual field on each side is divided by a line passing vertically through the fixation point (p. 1182) into two unequal parts, an outer or temporal and an inner or nasal. The latter is smaller owing to the shadow of the nose, its diameter being about 60° whereas the diameter of the temporal part is around 100°. Similarly a line passing horizontally through the point of fixation divides the field into an upper and a lower part, the former being restricted by from 5° to 10° by the upper lid and orbital margin. Rays of light from the outer or temporal half of the visual field fall upon the nasal (inner) half of the retina, those from the inner or nasal half of the visual field fall upon the temporal half of the retina (fig. 76 7, p. 1173). Although an image is formed upon each retina the two are fused in consciousness into a single impression (see p. 1178). In most animals the visual fields of the two eyes overlap, that is, certain parts of the outside world are seen by both eyes at the same instant—binocular vision. In animals with eyes placed laterally in the head overlap of the visual fields must obviously be very small in extent, the visual fields being almost completely separate—monocular vision. The extent of overlap of the visual fields of the monkey and of man, whose eyes are placed in the front of the head, is large (120° horizontal diameter) and the monocular field of vision, that is, the field which can be seen by one eye but not by the other, is relatively small (35°) (figs. 76 1 and 76 5). Rays

of light entering the eyes from an object in the binocular field of vision fall upon the nasal half of one retina and upon the temporal half of the other. If, however, the object is well to the right or to the left of the line of vision, i.e., in the outer part of either temporal field, the rays then fall upon the nasal half of the peripheral retina of the nearer eye (right or left) but not upon the other retina which is shaded by the nose. So the monocular field of vision consists in man of a crescentic area (35°) at the outer limit of the temporal field of each eye.

When one looks directly at an object the eyes are turned so as to bring an image of the object upon the most sensitive area of each retina, i.e., upon the nasal and temporal halves, respectively, of the foveae. The horizontal diameter of the entire visual field, that is, of the area of overlap of the two fields plus the monocular fields on each side is about 200°.

We shall see when the arrangement of the fibers conveying visual impulses is considered that fibers arising from the nasal halves of the retinas cross in the chiasma, whereas the temporal fibers remain uncrossed.¹ Thus it is that the occipital cortex of one side receives impulses from the nasal half of the opposite retina and from the temporal half of the retina of the same side, that is, those retinal halves which receive impressions from the opposite halves of the visual fields (fig. 76 7, p. 1173). Loss

¹ Thinking teleologically it is difficult to see a reason for the crossing of the optic fibers at the chiasma. But Adrian has pointed out that the natural course of the fibers is that which actually obtains, those from the right halves of the retina going to the right side of the brain, and those from the left halves to the left hemisphere. It is as though the homonymous retinal halves constituted separate organs. It is also suggested that the crossing of the pyramidal and the sensory tracts has been dictated by this arrangement of the optic fibers. For example, an object toward the right or the left side of the body will form an image on the left or the right halves, respectively, of the eyes. A movement of the head and limbs toward the object will entail contractions of muscles on that side. It is the most natural and convenient arrangement that the neurons governing such movements have their origins in the hemisphere that receives the visual impulses, and that sensory impulses initiated, as by touching the object, are returned to this side of the brain. It would be very confusing if visual impulses were received in one hemisphere and tactile impulses in the other.

of vision in one half of each eye is called hemianopia² (half-blindness). When the blindness affects the right or the left halves of both retinas, i.e., the nasal half of the left retina (temporal half of visual field) and the temporal half of the right (nasal half

or in both nasal retinal halves (temporal halves of visual fields) the hemianopia is crossed or heteronymous and is referred to as binasal or bi-temporal, respectively. It will be noted that the qualifying terms refer to the affected halves of the visual fields and not to the retinal halves. In other instances a quarter only of each visual field is affected, when the term quadrantic hemianopia is employed. It may be the upper or the lower quadrants of the nasal or of the temporal fields which are involved, or homonymous quadrants may be affected, i.e., a quadrant in the nasal field of one eye and in the temporal field of the other. The type of hemianopia is further specified by the use of the word "superior" or "inferior" (fig 76.3). Thus, a superior nasal quadrantic hemianopia is one in which the eyes do not see objects in the upper nasal quadrants of the visual fields (blindness of lower temporal retinal quadrants)

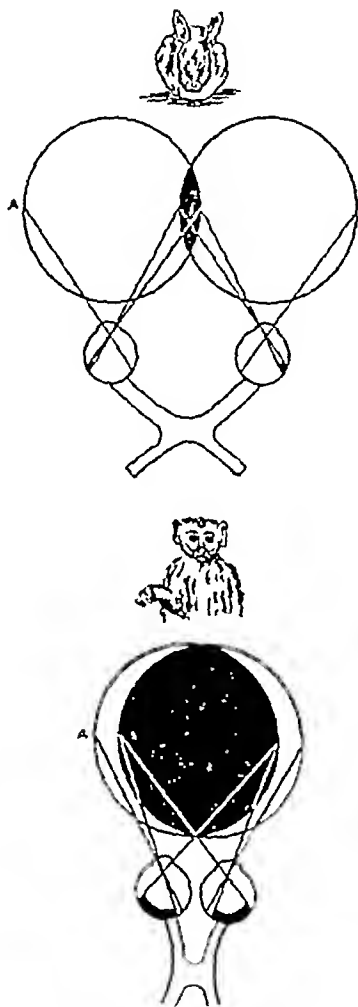


FIG 76.1 Showing monocular (white) and binocular (black) fields of vision. Upper rabbit, lower monkey (or man) (from Parsons, after Brouwer and Zeeman)

of the visual field) or vice versa, the hemianopia is said to be homonymous—left or right respectively. If the blindness is in the left half of one retina and in the right half of the other, i.e., in either both temporal retinal halves (nasal halves of visual fields)

² Hemianopsia, hemiopia (half sight) and hemiopia are alternative terms which are sometimes used, they have practically identical meanings

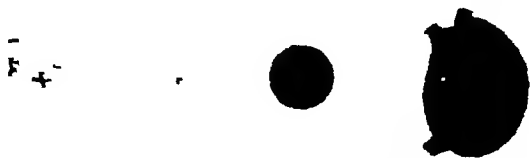


FIG 76.2 The blind spot. Close the left eye, hold the figure about six inches in front of the right eye and look steadily at the cross. Move the book slowly toward the eye until the circle disappears. When this occurs the image of the circle has fallen upon the entrance of the optic nerve from which rods and cones are absent, it is therefore insensitive to light. Figure on the right shows the blind spot projected 6° in front of the right eye as mapped out by means of perimetry (After Helmholtz)

Isopters: The sensitivity of the retina from the periphery to the fovea can be explored by using a series of test objects of graduated sizes with the perimeter, and determining in different meridians the distance from the fixation point (p 1182) at which each object is just perceptible. The points so determined are marked upon the perimeter chart and lines drawn through each set. Thus a series of boundary lines roughly concentric with the fovea are constructed which indicate the thresholds for the perception (minimum visible, p 1117) of the different test objects, and therefore demarcate levels of retinal sensitivity. Each is called an isopter of sensitivity, they might be compared to the contour lines indicating elevations on a detailed geographical map.

THE BLIND SPOT: The entrance of the optic nerve since it is devoid of rods and cones is completely blind. An object whose image falls upon it is therefore invisible (see fig 76.2). Not even a sensation of blackness results, for when the eye is fixated upon the cross, as

described in the legend of the figure, and the book moved until the optic disc is occupied by the circle the latter simply disappears, no sensation whatever being experienced to indicate its existence. Ordinarily this "hole" in the visual field causes no inconvenience because, in any position of the eyes, should one image of an object fall upon the blind spot a sensitive part of the opposite retina receives the other image. But even when one eye is closed the fine involuntary ocular movements, by constantly shifting the fixation point, though ever so slightly, keeps us completely unaware of the existence of the insensitive area. When mapped out accurately by perimetry the blind spot has the shape shown in figure 76.2

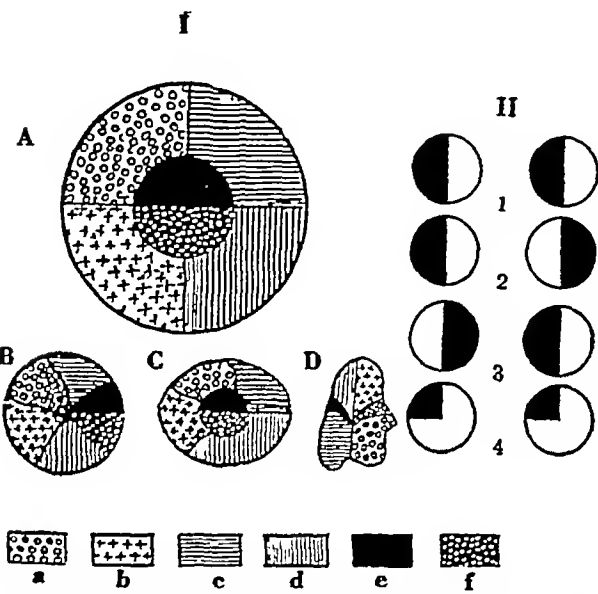


FIG 76.3 I, distribution of visual fibers A, right retina, B, optic nerve, C, optic chiasma, D, opposite lateral geniculate body, a, upper nasal quadrant of retina, b, lower nasal quadrant, c, upper temporal quadrant, d, lower temporal quadrant, e, upper macular fibers, f, lower macular fibers (Redrawn and modified from Henschen and Brouwer and Zeeman) II, Diagram of types of hemianopia, 1, homonymous, 2, bitemporal, 3, binasal, 4, homonymous superior quadrantic

Perimetry is the term applied to the procedure of mapping out the visual fields. The instrument employed is called a *perimeter*. It comprises a metal band or arm, shaped in a large arc of a circle with its concavity directed towards the subject (fig 76.4). A holder sliding in the arc carries the test object which can be moved centrally or peripherally as required. The arm itself is pivoted at the center enabling it to be rotated to any angle. The subject's head is supported on a chin rest. One eye is covered, the eye under examination is placed at the center of the sphere of which the perimeter arm forms the arc, and made to fix a point straight ahead in the center of the arm. The latter is rotated by degrees through a full circle and at each new position the test object is moved centrally until it is just

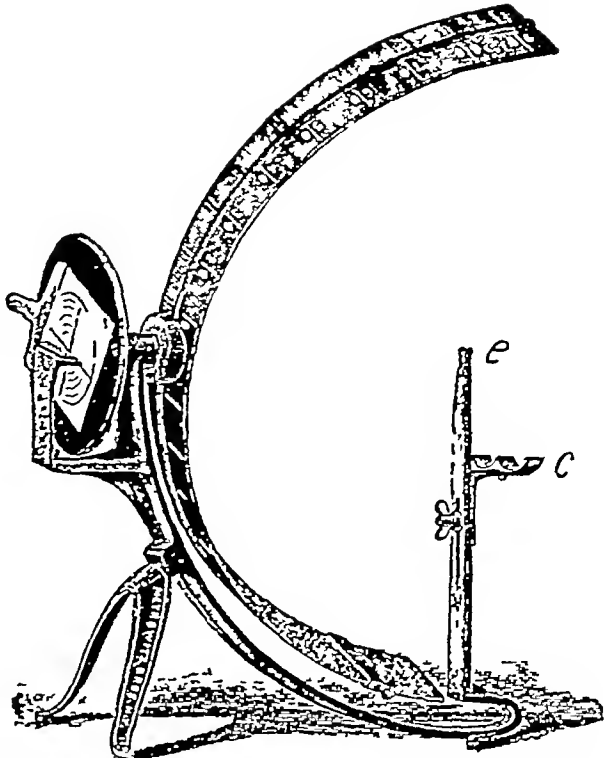


FIG 76.4 Perimeter, e, position of the patient's eye, c, chin-rest, further description in the text.

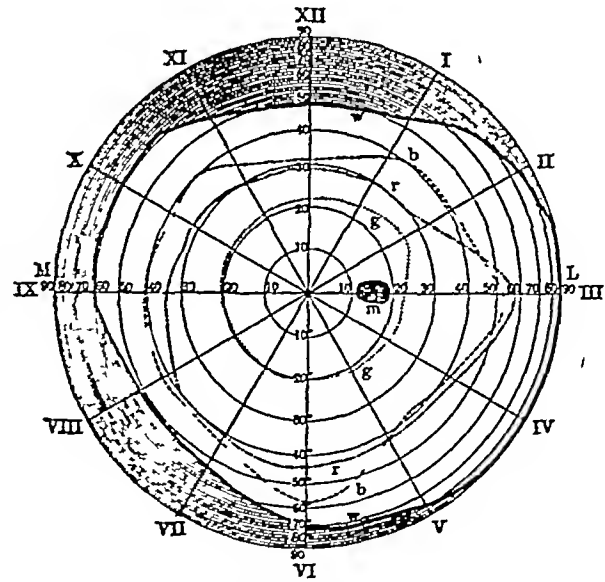


FIG 76.5 Perimeter chart showing normal visual field of right eye M, blind spot, g, r, b and w indicate boundaries of fields for green, red, blue and white, respectively Meridians indicated by Roman numerals (After Starling)

perceived by the subject. This point and corresponding points at the various positions of the arm are marked upon a perimeter chart and the contour of the visual field outlined through them. The chart (fig 76.5) is ruled in circles (comparable to latitudes on

a geographical map) to indicate degrees from the point of fixation, and in radiating lines ("longitudes" or meridians). The mapping of the blind spot and of the sensitivity of the retina in isopters has been referred to. The reader should consult texts on ophthalmology for a more detailed description of perimetry. A simple but rough method of perimetry (*confrontation method*) which will reveal a major limitation of the visual fields is the following. The observer stands facing the subject and about two feet in front of him. One eye of the patient is covered while the other, which he fixes upon the opposite eye of the examiner, is being tested. The examiner holds his finger midway between himself and the patient but outside the limit of his own visual field and then brings it slowly toward the midline. The observer compares the position at which he first sights his finger with that at which it is first seen by the patient. The procedure is repeated from various directions, above, below and from either side.

THE VISUAL PATHWAY

The optic nerve The ganglion cells of the retina (ganglion cell layer) whose central processes constitute the optic nerve, are the secondary neurons of the visual pathway (p. 1109). The primary neurons are the bipolar cells of the inner nuclear layer, the peripheral processes of the latter connecting with the visual receptors—the rods and cones. Tertiary neurons complete the pathway from the lateral geniculate body to the visual cortex. The fibers of the ganglion cells as they enter into the formation of the optic nerve are arranged in groups corresponding, in their relative positions, to the quadrants of the peripheral retina from which they arise. That is, fibers from the upper and lower temporal quadrants of the retina are found in the upper lateral and lower lateral regions of the nerve, respectively (fig. 76.3). Those from the nasal quadrants are situated in the inner sections of the nerve, the fibers from the upper quadrant lying above those from the lower. In the distal part of the nerve the macular or foveal fibers are in a lateral position, being wedged between the upper and lower temporal bundles, but again, fibers from the upper half of the fovea lie above those from the lower half.⁴

Throughout the visual pathway, this rather precise point to point localization of fibers from different retinal areas obtains.

⁴ There has been some dispute with respect to the existence of centrifugal fibers in the optic nerve. Such fibers are not present apparently in the optic nerve of higher animals, but they have been demonstrated in some cold blooded forms. Their function is unknown though it has been suggested that they control the motility of the pigment cells.

The optic chiasma The fibers from the temporal halves of the peripheral retina continue uncrossed into the corresponding optic tracts, while the fibers from the nasal halves cross in the chiasma to the optic tract of the opposite side.⁴ In the chiasma the temporal fibers lie in its lateral angle. Those of upper retinal origin lie above those arising from the lower retinal quadrant. In the crossing of the nasal fibers, those from the upper quadrants lie upon the upper aspect, those from the inferior quadrants upon the under aspect of the chiasma, the macular fibers lie in between.

The optic tracts and primary optic centers

The visual pathway from the chiasma to the primary optic center consists of compact bundles of fibers—the *optic tract*. In the optic tract, fibers from the upper quadrants of the peripheral retina tend to lie ventro-laterally, those from the lower quadrants ventro-medially. The foveal fibers lie dorso-laterally. The optic tract passes backwards and outwards between the tuber cinereum and the anterior perforated substance to the cerebral peduncle around which it turns as a flattened band to reach the *lateral (external) geniculate body*.⁵ In this, the *primary visual center*, the great majority

⁴ The proportion of the fibers which cross in the chiasma varies in different mammalian forms. In the opossum, example $\frac{1}{3}$ cross to the opposite optic tract, whereas in the ferret $\frac{2}{3}$ cross, in the rabbit also the uncrossed fibers are few in number. In the monkey as in man, all fibers from the temporal side of the retina including the corresponding half of the macula are uncrossed, the nasal side and the nasal half of the macula decussate with those of the opposite side. In submammalian forms all optic fibers cross.

⁵ The projection of the retina upon the external geniculate body has been studied by Brouwer and Zeman in monkeys. These observers produced localized lesions in the retina and examined the geniculate body histologically after time had been allowed for degeneration of the optic fibers. It was found that the retinas were not projected diffusely throughout the primary optic center, but showed localization of their different regions to definite sections. Thus the upper parts of the peripheral retinas, both nasal and temporal, of the two eyes were projected to the medial parts of the geniculate body. The lower parts of the retinas were represented in the lateral part (fig. 76.3, E). Dorsal to these regions and wedged between them was an extensive area wherein the macular fibers terminated. Binocular and monocular types of vision have separate representations. In the monkey the area which receives fibers from only one eye is very small, occupying a small rim on the ventral aspect of the primary center. In the rabbit, on the other hand, the monocular projection occupies almost the entire geniculate body, binocular vision being represented by a narrow area on the medial aspect of the geniculate body.

The cells of the lateral geniculate body with which the visual fibers synapse are disposed in six well defined

of the optic fibers make connection. A smaller number are continued to the *superior colliculus* (superior corpus quadrigeminum) but none of these are of foveal origin. The pulvinar of the optic thalamus has been generally looked upon as a relay station for visual impulses, but the researches of Brouwer and Zeeman lend no support to this belief, in monkeys, no degenerated fibers were observed to end in the pulvinar after excision of the eyeball. It has also been stated by Henschen that a lesion of the pulvinar in man does not cause hemianopia, whereas disease involving the lateral geniculate body, but which spares the pulvinar, produces complete hemianopia. The pulvinar may, however, be concerned with eye movements and visual judgments, e.g., the recognition of depth (stereoscopic function) and distance. Its connection with the angular gyrus which is a center for eye movements suggests that it serves these functions, but in order to do so it would require to be in communication also with the visual area of the cortex or with the external geniculate body. Such connections have not been demonstrated.

The superior colliculi. The superior colliculi are the chief centers for visual reflexes, and constitute as well the highest visual centers in submammalian forms. In the latter they are very prominent structures, in fishes and birds they are as large or larger than the cerebral hemispheres. But in higher animals, they are much reduced in size, and in man, are relatively inconspicuous structures, their visual functions having been taken over by the cerebral cortex.⁶ The superior colliculi of mammals receive afferent impulses from (a) the retina through the *optic tracts*, (b) the *occipital cortex* (visual area) and (c) the *spinal cord* via the spinotectal tract. Their efferent connection are with the medulla and the spinal cord through the *tectobulbar* and *tectospinal tracts*.

The reflexes centered in the superior colliculi serve to correlate eye movements with the movements of the head, or with the trunk and limbs.

laminae which have been numbered 1 to 6 from the surface inwards by Le Gros Clark. One set of three (1, 4 and 6) receive crossed fibers, i.e., fibers from the nasal half of the opposite retina. The other set (2, 3 and 5) are cell stations for the uncrossed fibers, i.e., from the temporal half of the retina of the same side.

⁶ Their visual function varies considerably in different classes of mammals. In the rat, cat, dog and other mammals of about the same phylogenetic level they serve as subcortical visual centers, crude visual perception being possible after complete removal of the visual area of the cortex. In man, bilateral destruction of the visual cortex causes total blindness.

The chief of these visual reflexes are, (a) turning the eyes in order to keep them fixed upon a stationary object when the head is turned in the opposite direction (compensatory reflexes), (b) movement of the head with the eyes so as to keep a moving object in view, and (c) movements of the limbs, neck or trunk as in avoiding a moving object or in fending off a blow threatening the eyes. Closure of the lids to protect the sight occurs simultaneously.

The optic radiation (geniculocalcarine pathway). The visual fibers after leaving the external geniculate body pass through the posterior extremity of the internal capsule, and curving forward and outward into the temporal lobe sweep backward in relation to the outer aspect of the posterior horn of the lateral ventricle to reach the *area striata* (area 17) of the occipital cortex (pp. 1031 and 1045). The optic radiation also contains descending fibers which end in the superior colliculus and the external geniculate body.

The visual cortex. This comprises that part of the cortex referred to above as the *area striata*,⁷ which forms the walls and lips of the calcarine fissure on the mesial aspect of the occipital lobe. The different retinal areas in their projections upon the cortex show definite localization. There is a point to point projection on to the striate area. The cortical cells, for example, are conceived as receiving individually impulses from single cones to form a pattern corresponding to that of retinal organization.⁸ The homonymous halves of the peripheral retinas are represented in the anterior part of the visual area, the upper quadrants in the upper wall and lip, the lower quadrants in the lower wall and lip. In other words, the nasal half of the right retina and the temporal half of the left are projected on to the left occipital cortex—the projection of the upper quadrant in each case lying above that of the lower quadrant. Similarly, the nasal half of the left retina and the temporal half of the right are projected on to the right striate area. The macular representation occupies the posterior part of the striate area reaching backwards to the occipital pole, but it also spreads forwards to overlap the projection area of

⁷ The histology of the striate area is described in ch. 68.

⁸ Though the existence of point to point projection of the retina upon the striate area has been definitely established, the cortical representation is not rigidly fixed and circumscribed by anatomical paths, but is capable of considerable functional adaptation and modification under changed conditions (see Holmes).

the peripheral retina (fig 766) As in the case of the peripheral retina, the upper and lower parts of the macula (fovea) are projected to the upper and lower halves, respectively, of this part of the striate area The cortical projection of the macula is possibly bilateral,⁹ i e, represented in its entirety in both hemispheres, since the retention of macular vision in both eyes after an extensive lesion of an occipital lobe is not infrequent Indeed sparing of macular vision has been reported after apparently

The subject of hemianopia may be quite unaware of his defect of vision He compensates sometimes to a considerable extent for his half-blindness by turning his eyes more toward the blind side of the visual field, so as to bring the image on to the sound part of the retina A new fovea (called a pseudofovea) is created in the peripheral retina of the seeing half of his eye which may become as sensitive as a normal fovea Compensation is also in part psychological, that portion

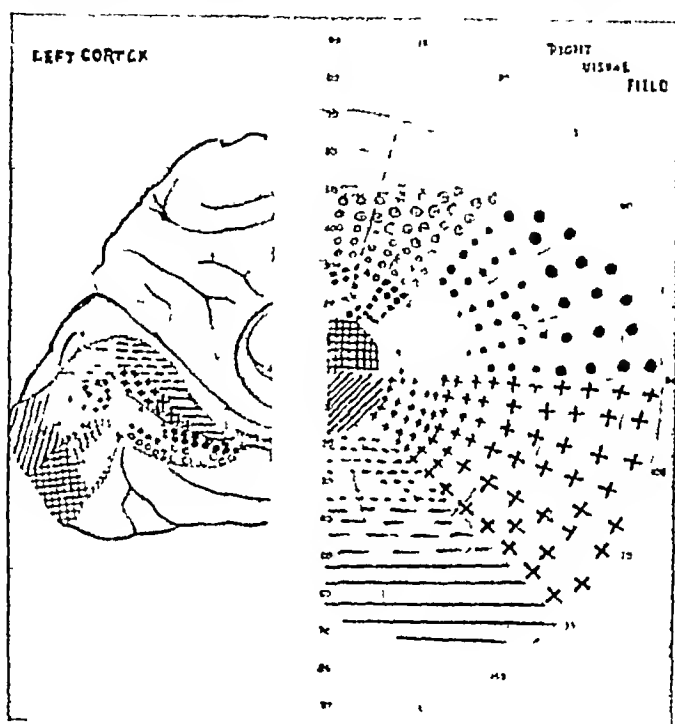


FIG 766 Diagram showing the projection of the retina on the calcarine cortex (from Holmes) Right hand figure is temporal half of right visual field

complete ablation of one occipital lobe It has been supposed by those who believe that the macula is represented bilaterally that fibers pass from the lateral geniculate body of one side through the posterior part of the corpus callosum to join the optic radiation of the opposite hemisphere In the monkey however excision of one occipital lobe is followed by retrograde degeneration of *all* cells of the corresponding lateral geniculate body, which would not be the case if some of its cells sent fibers to the opposite hemisphere In so far as man is concerned the question is unsettled

⁹ For views on this question see Penfield and associates, *Arch. Neur. Psych.*, 1935, 33, 816, and Maison and associates, *ibid*, 1938, 40, 981

of the image of a familiar object which falls upon the blind half of the retina is visualized mentally Thus the image is completed subjectively Normally, when an object is seen with one eye that part of the image which falls upon the blind spot is filled in mentally so that no gap is apparent

The effects of lesions at different levels of the visual pathway (see also p 1177)

(1) A destructive lesion of one optic nerve will result in total blindness of the corresponding eye (fig 767) Increased intracranial pressure may cause atrophy of the optic nerves and a gradual concentric reduction of the visual fields of both eyes

(2) A lesion involving the chiasma will result in visual defects whose nature will depend upon the fibers destroyed (a) Pressure upon the uncrossed fibers in the outer angle of the chiasma, as by an aneurysmal dilatation of the internal carotid artery, may produce blindness in the temporal part of the retina of the same side. If these fibers

from the lower retinal quadrants lie ventral to those from the upper, the lower are likely to be involved first in pituitary tumors, *superior temporal quadrantic hemianopia* will result. Lesions (e.g., tumors) pressing from above tend first to cause defects in the lower temporal quadrants of the visual fields. Dilatation of the third ventricle or a

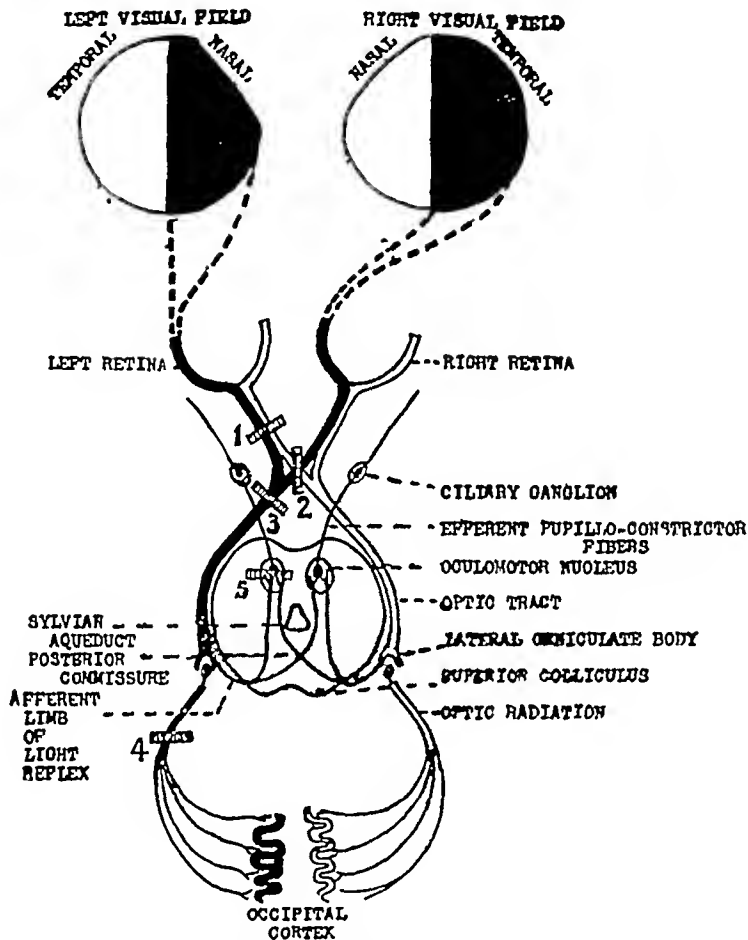


FIG 767 Diagram to show the effects upon vision and pupillary reactions resulting from interruption of retinal impulses at various levels 1, Optic nerve, blindness of corresponding eye, direct reaction of this eye and consensual reaction of sound eye lost. Sensual reaction of blind eye and direct reaction of sound eye retained. Near reflex unaffected. 2, Chiasma, bitemporal hemianopia. Wernicke's pupillary reaction. 3, Optic tract, homonymous hemianopia (blindness in nasal half of right retina and temporal half of left). Wernicke's pupillary reaction. 4, Optic radiation, homonymous hemianopia, light and near reflexes retained. 5, At synapses in the oculomotor nucleus. Light reflex lost, near reflex retained (Argyll-Robertson pupil, see also p. 1177). The lesion is usually bilateral (as in tabes, disseminated sclerosis, etc.)

on both sides are affected, the sight in the temporal half of each retina (nasal half of the visual field) may be lost (*bi-nasal hemianopia*). According to Cushing a dilated third ventricle may, by pressing from above, force the angles of the chiasma against sclerosed internal carotid arteries and so produce a bi-nasal hemianopia. (b) Pituitary tumors, owing to their position, are likely to involve the nasal fibers at the point of their crossing and thus cause *bi-temporal hemianopia*. Since the nasal

tumor of the pituitary stalk may produce such an effect.

(3) Lesions of the optic tract, of the primary visual center or of the optic radiation will result in *homonymous hemianopia*. The right halves of the two eyes (left halves of the visual fields) being effected in right-sided lesions and the left halves in left-sided lesions. An abscess or tumor of the temporal lobe may, by involving the optic tract or optic radiation, cause a homonymous hemianopia.

When the optic radiation is pressed upon by a temporal lobe lesion the hemianopia is very often incomplete, i.e., quadrantic. The ventral fibers of the radiation are likely to be implicated by a tumor in the lower part of the lobe, and a superior quadrantic hemianopia result. Injury to the dorsal fibers tends to cause a defect confined at first to the lower homonymous quarters of the visual fields.

(4) *Lesions of the occipital cortex.* A lesion involving the area striata of one hemisphere or the optic radiation before their termination therein results in an homonymous hemianopia, right or left, depending upon the side of the brain affected. Quadrantic homonymous hemianopia will result when the lesion is restricted to the upper or lower part of the striate area. Owing to the large cortical area representing the macular region, or perhaps to the fact that the macula is bilaterally represented, extensive unilateral occipital lesions or excision of a large part of this area often leave acute (central) vision intact, the blindness then involving only the peripheral half of each retina.

THE PUPIL

Pupillary reflexes. Reflex changes in the size of the pupil occur under the following conditions: (a) Constriction or dilatation occurs in response to changes in light intensity (*light reflex*). When light is thrown into the normal eye the pupil of that eye constricts promptly, this is the *direct pupillary reaction*. But the pupil of the opposite eye, though shaded, also narrows, this, the *indirect or consensual pupillary reaction*, is dependent upon fibers which cross to the pupillary-constrictor center of the opposite side. (b) Constriction occurs as a part of the mechanism of accommodation to near vision (*pupillary reaction of accommodation*). With it are associated convergence of the eyes and accommodation of the lens. These three reactions are appropriately grouped under the term *near reflex* or *accommodation reflex*. (c) Dilatation follows stimulation of the skin of the neck (*carotid-spinal reflex*). (d) Irritation of the eyelid or conjunctiva causes dilatation followed by constriction of both pupils (*oculo-sensory reflex*). The afferent fibers of this reflex are contained in the ophthalmic division of the trigeminal nerve. (e) Closure of the eyelid, or an effort made to close the lid while it is forcibly held open causes slight pupillary constriction (*lid or orbicularis reflex*). (f) Pupillary dilatation may occur during certain *emotional states* (e.g., fear), as a result of acute pain or a sudden

sound. The pupil is extremely sensitive to the latter type of stimulus, responding by dilatation to sounds of different pitches at the lowest intensities which can be heard. (e) Finally, stimulation of labyrinthine receptors causes changes in the diameter of the pupil (p. 982). For example, rapid rotation of the body around its long axis causes dilatation of the pupil and rhythmical changes in its diameter (*labyrinth*).

THE IRIS

The iris is the most anterior part of the vascular tunic of the eye. It is a thin contractile disc perforated a little to the nasal side of its center by the pupil. The pupillary margin rests upon the anterior surface of the lens. The space between the lens and the cornea is divided by the iris into a larger *anterior* and smaller *posterior chamber*, the two chambers communicating through the pupil. The periphery (root) of the iris is attached to the anterior surface of the ciliary body and is continuous through the pectinate ligament (p. 1165) with the posterior elastic lamina of the cornea. The following five layers from before backwards compose the structure of the iris, (a) the *anterior epithelium*, (b) the *anterior limiting membrane*, (c) the *stroma*, (d) the *posterior membrane* and (e) the *posterior epithelium*. The *anterior epithelium* consists of a single layer of flat endothelial like cells. Near the pupillary margin of the iris there are many small pits—the *crypts of Fuchs*—over which the epithelium is absent. The stroma is composed of loose connective tissue. It transmits the vessels and nerves and holds numerous branched cells which in dark eyes contain pigment granules. The iris contains two involuntary muscles—the *sphincter pupillae* and the *dilator pupillae*. The *sphincter pupillae* is embedded in the stroma and comprises a band of circular fibers about 1 mm. broad surrounding the pupil. When these fibers contract the pupil is constricted. The *dilator pupillae* constitutes the fourth layer of the iris, i.e., the posterior membrane, mentioned above. It consists of a thin layer of smooth muscle fibers which converge towards the pupillary margin where they blend with the fibers of the sphincter. At the root of the iris the dilator fibers pass into the ciliary body from which they take origin, when they contract they draw upon the pupillary margin and thus dilate the pupil. The *posterior epithelium* comprises two layers of deeply pigmented cubical cells, it is the continuation anteriorly of the pars ciliaris retinae (p. 1152). The arteries of the iris which are loosely coiled form two vascular circles, one near the pupillary margin—the *circulus arteriosus minor*, the other near the root of the iris—the *circulus arteriosus major*. The two circles are connected by vessels which, arising from the larger circle, converge towards the pupillary margin where they form the smaller circle.

Blue or gray eyes owe their appearance to the pig-

ment in the posterior epithelial layer as seen through the unpigmented stroma and other layers of the iris. The pigment cells of the stroma are responsible for the color of dark eyes, the shade varying with the quantity of pigment present. In the white races nearly all newborn babies have blue eyes because pigment does not develop in the stroma until some weeks after birth. But Negro babies and others belonging to the dark races have brown eyes, the stroma pigment being well developed at birth.

The functions of the iris The iris has three important functions, (a) it serves as an opaque screen, like the diaphragm or "stop" of a camera, to adjust the quantity of light reaching the retina under different intensities of illumination, (b) it prevents light from passing through the periphery of the lens and thus reduces spherical and chromatic aberration. The image is thus more clearly defined by restricting the transmission of light through the central part of the lens, and (c) when the pupil constricts the depth of focus of the eye is increased (p 1156).

Pupillo-constrictor pathways

It is generally believed that the receptors of the light reflex are the same as those mediating visual sensations. Wagman and his associates have made the interesting observation that in dim light the curve of pupillary size at different wave-lengths agrees closely with the scotopic luminosity curve. The most effective wave-length in causing pupillary constriction of the dark adapted eye is, therefore, around 510 $m\mu$, in the light adapted eye, it is at about 560 $m\mu$. Thus, the pupillary reaction shows a difference in sensitivity at high and low light intensities corresponding to the Purkinje shift.

The afferent fibers of the light reflex travel with the visual fibers and with the afferent fibers of the dilator response as far as the lateral geniculate bodies. Here they part company from the latter two sets of fibers. They do not enter the lateral geniculate body, but pass into the brachium of the superior colliculus. They then proceed to a group of cells in the pretectal region—the *pretectal nucleus*—where they make their first synaptic contacts. The impulses are finally conveyed by secondary neurons to the oculomotor nucleus on both sides of the brain, but mostly to that of the opposite side. The superior colliculus itself is not interposed in their path, for Ranson and Magoun did not observe pupillary constriction when this part was stimulated. The partial decussation of the

afferent fibers, occurs, Ranson and Magoun believe, in the posterior commissure. An earlier crossing of some fibers occurs also in the optic chiasma. The fibers to the same side pass caudally and ventrally along the side of the central gray matter. Though under ordinary circumstances reduction in the tone of the pupillo-dilator center occurs reciprocally with activation of the pupillo-constrictor center, the light reflex can be carried out through the latter alone. The reflex is therefore retained after section of the cervical sympathetic (which conveys the dilator fibers). Since the afferent fibers mediating the light reflex separate from the visual pathway at the level of the lateral geniculate body, lesions of the optic pathway beyond this point do not interfere with the light reflex.

The *efferent* fibers subserving the light reflex belong to the parasympathetic division of the autonomic nervous system. They originate in the oculomotor nucleus (probably the Edinger-Westphal nucleus) and are conveyed to the iris (sphincter pupillae) via the third nerve, ciliary ganglion and short ciliary nerves (fig 76 8). The *near reflex* is dependent upon cortical centers. Impulses pass by association fibers from the occipital to the frontal cortex (frontal eye field) and thence via the internal capsule to the nucleus of the 3rd nerve. Constriction of the pupil which accompanies accommodation of the lens is brought about through fibers which probably pass directly to the pretectal region from the occipital cortex adjacent to the visual area. The efferent path from the oculomotor nucleus is the same as that for the light reflex. The afferent pathway is via the visual fibers, i.e., lateral geniculate body and optic radiation, not through the superior colliculus. The reflex is bilateral, i.e., it occurs in both eyes when one is covered and the other directed to a near object.

The pupillo-dilator pathway

The dilator muscle of the pupil receives sympathetic fibers which arise from the 1st and 2nd thoracic segments of the spinal cord, and sometimes from the 8th cervical or the 3rd thoracic. They issue by the white rami and pass via the cervical sympathetic to the superior cervical ganglion. From here postganglionic fibers are conveyed along the internal carotid artery into the cranial cavity. Entering the trunk of the nasociliary branch of the first division of the fifth nerve, they are transmitted to the iris in the long ciliary nerves. Some fibers also pass without interruption through

the ciliary ganglion into the short ciliary nerves. These fibers supply in addition the smooth muscle of the orbit which lies in relation to the capsule of Tenon (fascia bulbi) and in the "check ligaments" of the ocular muscles. The smooth muscle forming the deep layer of the levator palpebrae superioris also receives innervation from the sympathetic.

terior commissure to enter the superior colliculus. The pupillo-dilator reflex involves reciprocal inhibition of pupillo-constrictor tone. This is the paramount factor in the dilator reflex, for after section of the sympathetic pupillary fibers, the pupil dilates in the dark, or in response to a painful stimulus or emotional excitement, in an almost normal fashion. Another factor in the dilatation of

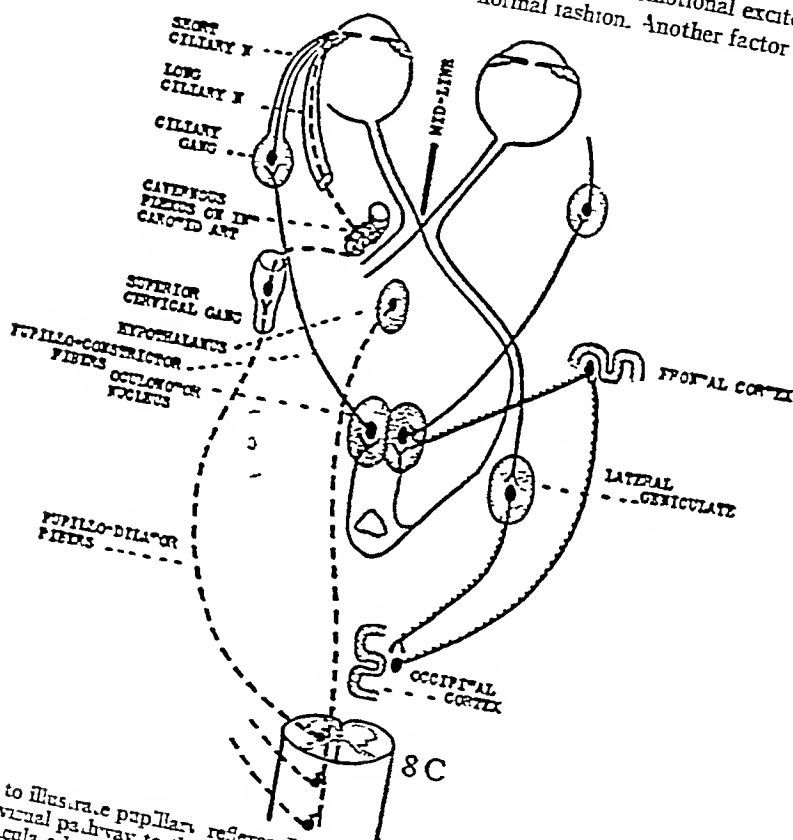


FIG 76S Diagram to illustrate pupillary reflexes. Interupted lines indicate the dilator reflex, plain lines the constrictor reflex and visual pathway to the level of the lateral geniculate body, hatched lines show path of near reflex from lateral geniculate body to the cortex and nucleus of the 3rd nerve.

The afferent pathway of the pupillo-dilator reflex accompanies the pupillo-constrictor fibers as far as the lateral geniculate body. From here on their course has not been clearly defined, but they presumably descend through the tegmentum of the m.d-brain, and the reticular formation of the pons, medulla and spinal cord to reach the cilio-

A higher pupillo-dilator center is situated in the hypothalamus, which, in turn, is probably connected with the cortex of the frontal lobe. It sends fibers through the m.d-brain which, according to the researches of Beattie, pass ventrally by the pos-

terior commissure to enter the superior colliculus. The pupillo-dilator reflex involves reciprocal inhibition of pupillo-constrictor tone. This is the paramount factor in the dilator reflex, for after section of the sympathetic pupillary fibers, the pupil dilates in the dark, or in response to a painful stimulus or emotional excitement, in an almost normal fashion. Another factor in the dilatation of

the pupil by sympathetic stimulation is the constriction of the vessels of the iris. Indeed it has been doubted that a dilator muscle really exists, the dilator action being thought to be the result of inhibition of constrictor tone and the vasoconstriction just mentioned.

When the sympathetic fibers are paralyzed the pupil is narrowed as a result of the unbalanced action of the constrictor fibers, and the dilatation of the pupil which normally follows the application of a stimulus, such as a scratch or pinch, to the skin of the neck (cilio-pinal reflex) fails to occur. The pupil, however, still reacts to light. Drooping

of the upper lid (ptosis) will result from paralysis of the smooth muscle of the levator palpebrae superioris and there may be recession of the eyeball (enophthalmos) from paralysis of the unstriated muscle of the orbit (see p 1105)

An investigation of the pupillary reactions may give valuable information concerning the site of a lesion in the brain (see fig 76 7, p 1173) (a) A lesion destroying one optic nerve, since it interrupts the afferent pathway but leaves the efferent intact, abolishes the direct but not the indirect (consensual) reaction on the blind side. The direct reaction on the sound side is, of course, retained but the indirect is lost. That is to say, a light thrown into the sound eye causes a response in this eye as well as in the blind eye, but a light thrown into the blind eye is without effect upon either eye. The near (accommodation) reflex is not abolished. Blindness due to destruction of both optic nerves results in the loss of the reflexes for light as well as of those for accommodation. (b) Hemianopia due to a lesion of the chiasma, or of the optic tract, results in the loss of both the direct and indirect reactions to light thrown upon the blind half of either retina. Light falling upon the sound halves of the retinas causes the normal response. This is known as *Wernicke's hemianopic pupillary reaction*¹⁰ (c) Loss of the light reflex (both direct and indirect reactions) with retention or even exaggeration of the accommodation-convergence reflex is known as the Argyll-Robertson pupil. The pupil is also, as a rule, smaller than normal (myosis) and does not dilate to a painful stimulus nor fully to atropinization, the vestibular reflex is frequently absent as well. The abnormal pupillary responses occur quite apart from any defect of vision. Though usually bilateral the Argyll-Robertson pupil is sometimes confined to one side. It is most commonly seen in syphilitic degeneration of the central nervous system (e g, tabes), but occurs occasionally in other conditions as well. The site of the lesion responsible for the Argyll-Robertson pupil is not known precisely. Merritt and Moore give evidence for placing it in the region of the posterior commissure where presumably, the neighboring dilator pathway is also interrupted, but the pathway for the accommodation reflex from the cortex to the oculomotor nucleus would be spared. Scala and Spiegel believe from the results of their experiments

¹⁰ Owing to the difficulty of confining a beam of light to the blind half of the retina it is not an easy matter to demonstrate this reaction

that the disease involves the synapses between the afferent and efferent neurons of the light reflex, that is, in the oculomotor nucleus itself. But the oculomotor nerve supplies both pupils and a lesion here is difficult to reconcile with the unilateral loss of the light reflex which sometimes occurs. Langworthy and Ortego conclude, after a careful study, that the lesion is peripheral—due to changes in the iris itself involving sympathetic, parasympathetic and sensory nerves as well as the muscle cells of the sphincter. This theory can account for the irregularity of the pupil, and perhaps, if the sympathetic innervation were injured in greater degree than the parasympathetic, for the myosis, but it cannot explain the preservation of the accommodation reflex. (d) Destruction of the oculomotor nucleus or of the efferent pathway abolishes all light and accommodation reactions on the same side. The direct and indirect reactions are retained on the contralateral side. (e) Lesions involving the visual pathway after the separation of the visual and pupillary fibers, e g, lateral geniculate body, optic radiations or occipital cortex, leave the light reflex unaffected. (f) A bilateral lesion implicating the pathway from the cortex to the center for accommodation in the oculomotor nucleus will cause a loss of the near reflex and leave the light reflexes intact. This is the converse of the Argyll-Robertson pupil, and is sometimes seen in post-diphtheritic paralysis.

The effects of drugs upon the pupil and ciliary muscle

Dilatation of the pupil is spoken of as *mydriasis*, constriction as *miosis*. Drugs which cause pupillary dilatation are therefore called *mydriatics*, those which cause constriction, *miotics*. Paralysis of the ciliary muscles is known as *cycloplegia*, drugs which cause this effect are called *cycloplegics*.

Mydriasis is caused by drugs which

(a) Paralyze the peripheral constrictor (parasympathetic) mechanism, such as atropine, or homatropine. Atropine is also cycloplegic, homatropine much less so.

(b) Stimulate the dilator (sympathetic) mechanism, e g, adrenaline, cocaine. These drugs have no effect upon accommodation.

Miosis is caused by drugs which

(a) Stimulate the peripheral constrictor mechanism, e g, pilocarpine, physostigmine, muscarine. These drugs also cause spasm of the ciliary muscles.

(b) Diminish the inhibition of the constrictor center, e g, morphine. The action of this drug upon the pupil depends largely upon the intensity of the illumination. It appears, therefore, to exert its effect, mainly, by increasing the sensitivity of the light reflex.

(c) Stimulate the constrictor center, e g, picrotoxin.

Corresponding retinal points—The horopter

When the gaze is directed to an object, an image is formed by each eye and impulses are conveyed to both sides of the brain, yet perfect fusion of the two images occurs in consciousness and only one image is seen. This characteristic of *binocular vision* is explained upon the theory of *corresponding retinal points*. The corresponding points in the retinas (foveas) which when stimulated simultaneously cause a single visual sensation, lie in the nasal half of one retina and the temporal half of the other. When the eyes are converged, the retinas are turned so that the images fall upon these corresponding parts. If, as a result of unequal

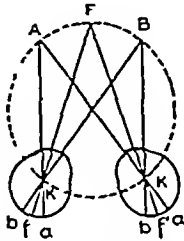


FIG. 769 The horopter (Müller). *F* is the fixation point. The images of *A*, *F*, and *B* fall upon the retinae at corresponding points as *aa'*, *ff'*, and *bb'*. The projection of all such corresponding points lies upon the circumference of the dotted circle. It is obvious that there is a different horopter for each position of *F*. (After Duke Elder, *Text Book of Ophthalmology*.)

action of the ocular muscles, this cannot be brought about, the separate images are not fused in consciousness and an object appears double. This abnormality of vision is known as *diplopia*.

The horopter. When the eyes are fixed upon a point in space, a number of other points can be located by calculation which are projected upon corresponding points of the two retinas (foveas). A line joining such points forms a circle called the *horopter* which passes through the fixation point and the nodal points of the eyes. The horopter will vary of course with the point of fixation of the eyes and does not exist unless the eyes act synergically (see fig. 769). Points in the visual field lying outside the horopter do not fall on the corresponding points in the two retinas (peripheral retinas) and, as a consequence, actually cause a double impression. But this *physiological diplopia* as it is called does not thrust itself upon consciousness, it is suppressed or ignored and therefore does not cause confusion. Yet one can easily

demonstrate for himself that it exists. For example, when the eyes are fixated upon a near object, such as a pencil tip held close to the face, a more distant object may, through a conscious effort, be observed in duplicate. For this reason, it is also sometimes referred to as *introspective diplopia*.

The double image is always projected to the plane of the object upon which the eyes are fixed, and the doubling of the image is either homonymous (uncrossed) or heteronymous (crossed) depending upon whether the object which produces the double image is beyond the point of fixation or between it and the eyes. That is to say, when the object (*D* in figure 76 10) whose images fall on non-corresponding retinal points is closer to the eyes than the object upon which the eyes are fixed, the images, *I* and *I*, are projected across the lines of sight, the right hand image being formed by the left eye and the left hand image by the right eye. When the object is beyond the plane of fixation the projection of the images (*I*₁ and *I*₂) is homonymous, each being formed by the corresponding eye. The reader may demonstrate these facts for himself. When a pencil is held close to the eyes so as to form a double image while the eyes are focussed upon a more distant object, closing the right eye causes the left hand image to disappear, whereas, if the eyes are focussed upon the pencil and the object giving rise to the double image is more distant, closure of the right eye abolishes the right hand image. Closure of the left eye, of course, produces converse effects.

DEPTH PERCEPTION STEREOSCOPIC VISION (GR. STEREOS, SOLID SKOPEO, I VIEW)

Our visual judgment of solidity, that is, our recognition that the object has depth as well as height and width, is due largely to the fact that vision is normally binocular and corresponding points in the two retinas receive slightly dissimilar images of any given object. If the reader will look at some object in front of him, first closing one eye and then the other, he will find that the view seen by the right eye is slightly different from that seen by the left (fig. 76 11). The right eye is able to see more of the right side of the object, the left eye more of the left side. The two slightly disparate images are fused in the brain, yet the composite image has hidden within it something of each separate one, upon this the stereoscopic effect to a large extent depends. The fusion of the dissimilar images by the brain, and the impression of depth

and solidity produced thereby, lies in a field of psychology of which little is known

In order for two dissimilar images to be fused in consciousness, it is not necessary that they fall

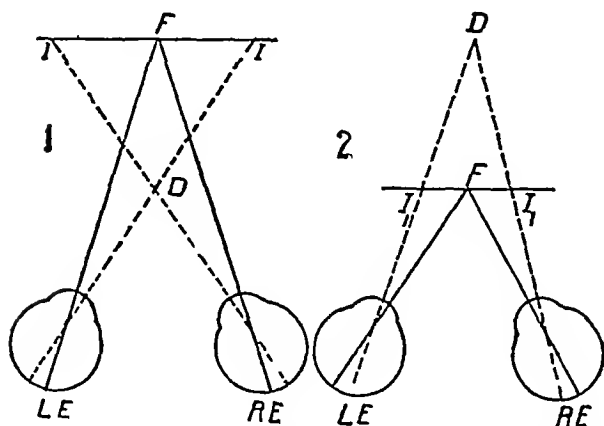


FIG 76 10 Illustrating the projection of retinal images that fall upon non-corresponding retinal points, i.e., on different sides of the foveae—physiological diplopia *I*, heteronymous diplopia (crossed), homonymous (uncrossed) diplopia, *D*, positions of object in relation to the fixation point. The images, *I*, *I* and *I*₁, *I*₁ are projected to the plane of the object *F* upon which the eyes are fixed, *LE*, left eye, *RE*, right eye

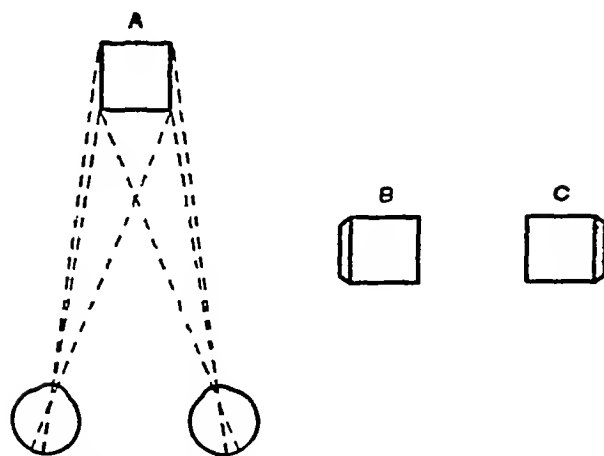


FIG 76 11 Examples of aspect disparity. *A* is a view, seen from above, of lines of sight when the eyes look at a solid object. Notice that the right eye sees some of the right hand side of the object, while the left eye sees some of the left hand side

B is a front view of a cube seen by the left eye with the right eye closed. *C* is a front view seen by the right eye with the left eye closed. When both eyes are open, we see a fusion of *B* and *C* (After Graham)

upon retinal points which correspond exactly, unification of the images results though there is some degree of non-correspondence. Actually, there is a greater impression of depth when the dissimilar images do not fall on retinal points which fail to correspond perfectly, provided that the discrep-

ancy is not so great as to prevent fusion. In figure 76 12 a scene of a tent and pine trees is represented diagrammatically. When the two scenes are fused the tent is projected closer or farther from the eyes according to its position in the two pictures in lateral relation to the pines. When in one or in both pictures it is moved toward the mid-line it appears to advance in front of the pine trees,

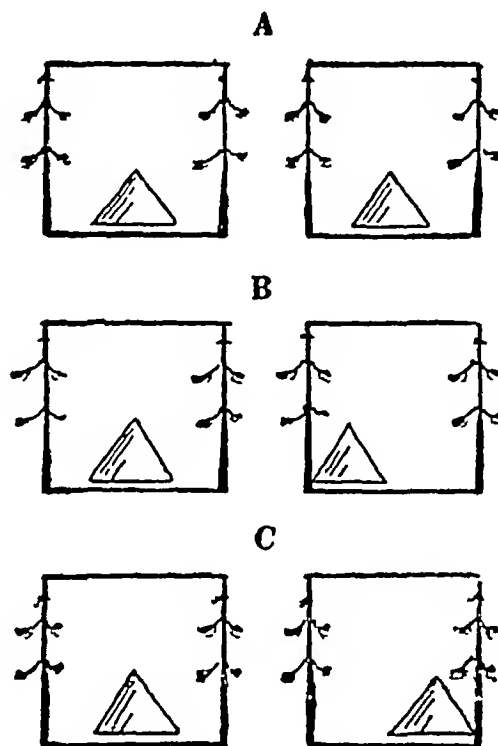


FIG 76 12 Scenes of a tent and pine trees (diagrammatic) To show the impression of depth caused by the fusion of two dissimilar images. If the figures be gazed at steadily at a distance of about 5 inches from the eyes and a card held edgewise from the tip of the nose to the line of letters each horizontal pair of figures can, with a little practice be made to fuse. In *A* the aspect of the two scenes is practically the same and the fused picture appears quite flat. In *B*, the tent in the right hand figure is placed more toward the mid-line and upon fusion appears to be in front of the pines. In *C*, the disparity of the right hand figure is farther from the mid line, the tent is now projected to a point behind the pines

when moved laterally it seems to recede. The three dimensional effect is enhanced by increasing the non-correspondence, as when the tent in one or both pictures is moved, to or away from the mid-line. But if moved too far the non-correspondence is too great to permit fusion to occur and the stereoscopic effect is lost.

The instrument known as a stereoscope produces an illusion of solidity by making use of the principle of simultaneous stimulation of the retinas

by dissimilar images. A photograph taken with an ordinary camera appears flat because identical images are formed upon the retinas. A stereoscopic photograph, on the other hand, is taken by a camera provided with two lenses which are set, like the eyes, a short distance apart. Thus, an illusion of depth is produced. Two slightly dissimilar views are taken which, when looked at through the stereoscope, are projected by means of prisms, one to each eye, so as to fall on corresponding retinal points (fig. 76 13). Depth perception cannot, however, be explained entirely upon the basis of dissimilar retinal images, for though the discrimination of depth (or distance) is much more acute in binocular vision it is not abolished when one eye is closed. For example, when two objects are placed one in front of the other and viewed binocularly from a distance of about 6 meters the least distance between them which can be perceived by average normal eyes is around 20 mm. When one eye is covered the least perceptible difference is increased to 120 mm. The ability to detect a small difference in distance from the eyes of two objects, and so to appreciate depth and solidity, is called *stereoscopic acuity*. It is expressed as the least difference between the angles formed by the lines of sight to two objects when one is just perceived to be farther away than the other (fig. 76 14). The difference in angles is known as *stereoscopic parallax*, and may, in a person with very high visual acuity, be as little as 2 seconds of arc.

Other factors, listed below, do not depend upon binocular vision, but play an important part in depth perception through one eye alone (monocular vision).

(1) *The apparent size of various objects in our field of vision*. We know from experience the approximate dimensions of the objects which we see, but the image which an object casts upon the fovea diminishes as its distance increases (p. 1118). For example, a church steeple at a distance casts an image upon the retina no larger, perhaps smaller, than would a pencil held a few inches from the eyes. We know the relative sizes of the two objects, and therefore infer that the steeple must be far away and the pencil near.

(2) *Accommodation of the eye*. Since a near object requires a greater effort of accommodation than does a more distant one for its image to be focussed upon the retina, some cue may possibly be given as to the relative distance of two objects from the eyes. This factor however is of very minor importance, and may be negligible.

(3) *The apparent change in color of an object with*

distance. The atmosphere is not perfectly transparent or equally so for all wave lengths. Tree-clad hills, which we know to be green, appear bluish in the distance, the colors of many other objects appear to fade with distance, their detail and outline being dimmed by haze.

(4) *The blocking out of parts of a distant view by objects between it and the eyes* gives a sensation of depth,

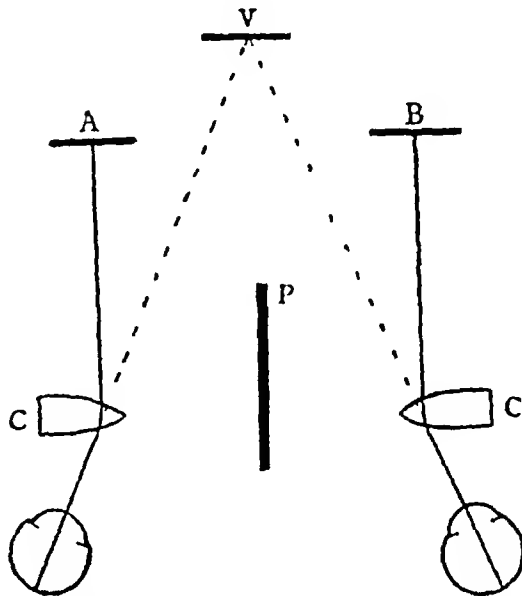


FIG. 76 13. Diagram to illustrate the principle of the stereoscope (Brewster's). A and B represent photographs of two scenes, slightly dissimilar because they were taken from different positions. C, curved prisms, p, partition to prevent one eye seeing the picture opposite the other eye, L.E., left eye, R.E., right eye. When the two pictures are viewed through the instrument, they are fused perfectly into one of apparently three dimensions, points in the background of which seem to be situated at V. A similar stereoscopic effect is produced when two scenes are printed each in a different color, e.g., red and blue, and imperfectly superimposed so their outlines overlap. If they are viewed in the ordinary way, they appear as a flat jumbled picture, but if a red glass be held in front of one eye and a blue glass in front of the other, so as to sort out the separate scenes, one to each eye, a clear black and white view is seen in three dimensions.

for the overlapping of parts of farther objects by nearer ones gives an indication of their relative distance from the eyes.

(5) *Linear perspective*. Straight lines running into the distance which are actually parallel (or objects along imaginary straight lines) are convergent in the retinal image. When we look down a railway track, for example, the rails appear to converge towards some point beyond the horizon. This arrangement of lines in the retinal image we have come to associate with distance. It depends upon the fact that points at a

constant distance apart subtend a smaller angle at the eye the farther they are removed. The artist draws objects along imaginary lines which run towards a point in the background of his picture.

(6) *Parallax* When one moves in any direction, near objects appear to move in the opposite direction, those in the background in the same direction as ourselves. This apparent movement of near objects in relation to ones farther away is called parallax, it is also produced by a movement of the head or eyes, even though the body remains stationary. Now, involuntary movements

muscle component of the elevator of the lid is innervated by the sympathetic.

THE NUCLEUS OF THE 3RD NERVE is situated in the floor of the Sylvian (cerebral) aqueduct and subjacent to the superior colliculus. It is in close relation to the medial longitudinal fasciculus. It is composed of a group of five smaller nuclei (fig. 76.15).

(a) The *central nucleus* (Perlia's nucleus) fuses with its fellow of the opposite side to form a single gray mass in the mid-line. It is probably the center for convergence of the eyes (internal recti). (b) The *caudal*

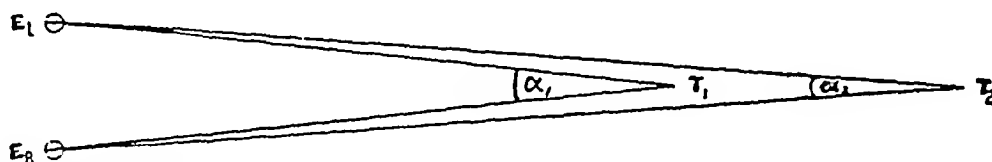


FIG. 76.14 Stereoscopic parallax in the viewing of two objects T_1 and T_2 . It is expressed as the difference between the angles α_1 and α_2 .

of the eyes are continually taking place with the production of parallax, this is believed to be an important but not an essential factor in giving us a sense of depth. That it is not an essential factor in depth perception was proved by Dove (1841) who found that objects illuminated by an electric spark were seen in three dimensions, the stereoscopic effect, as was shown later by Volkman, is experienced, though the duration of the flash is only 0.000,001 second.

(7) The distribution of light and shade over the surface of an object and the shadow which it casts upon its surroundings is also an important factor in the production of the stereoscopic effect.

Normally, the images formed by the two eyes are very nearly equal in size, varying by less than one per cent. When they differ in size to a degree which prevents perfect fusion with a consequent impairment of binocular vision the condition is spoken of as *aniseikonia*. Little interference with binocular vision results unless the inequality of the images is more than four or five per cent. Aniseikonia may be a factor, according to Bielschowsky, in the production of strabismus.

OCULAR MOVEMENTS

The innervation of the ocular muscles

The nerves supplying the extrinsic muscles of the eye are the 3rd (oculomotor), 4th (trochlear) and the 6th (abducent). The oculomotor nerve supplies all the extrinsic muscles of the eyeball except the superior oblique and the external rectus. It also supplies the striated portion of the levator palpebrae superioris and conveys parasympathetic fibers to the sphincter pupillae (p. 1175) and ciliary muscle. The deep smooth

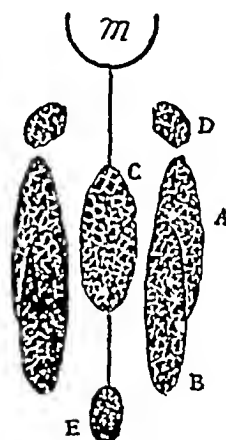


FIG. 76.15 A scheme of the various groups of cells which together constitute the nucleus of the oculomotor nerve: A, the dorsolateral nucleus, B, the ventromedial nucleus, C, the central nucleus, D, the Edinger-Westphal nucleus, E, the caudal central nucleus, M, the third ventricle (from Gray, *Anatomy of the Human Body*, after le Gros Clark, redrawn).

central nucleus lies in line with and behind the former. It also fuses with its fellow of the opposite side. Functionally it is considered a part of the central nucleus and is shown by Brouwer and others as actually continuous with the latter. (c) The *dorsolateral nucleus*. It and the next two nuclei are paired. The dorsolateral nucleus is probably the center for upward movements of the eyes (superior rectus and inferior oblique muscles). The *striped* muscle of the levator palpebrae superioris also, it is believed, receives its innervation from this nucleus. (d) The *ventromedial nucleus*, lying medial, ventral and caudad to the preceding is thought to be concerned with downward movements (inferior rectus). (e) The *Edinger-Westphal nucleus* lies on each side dorsal and lateral to the central nucleus. It is

believed to supply fibers to the sphincter pupillae and ciliary muscle

The axons arising from these cell groups pass for the most part into the nerve of the same side. A few fibers supplying the internal and inferior recti and inferior oblique muscles decussate with those of the opposite side.

The fibers after issuing from the oculomotor nucleus form a well-defined tract (tract of the oculomotor nerve) which runs downwards and forwards through the tegmentum, traversing the red nucleus and medial portion of the substantia nigra. They emerge from the medial aspect of the cerebral peduncle.

THE NUCLEUS OF THE TROCHLEAR NERVE lies in the floor of the cerebral aqueduct adjacent to the posterior end of the ventral medial nucleus of the oculomotor nerve, and on a level with the inferior colliculus. It supplies the superior oblique, and, with the ventral-medial part of the oculomotor nucleus, forms a center for downward movements of the eye. The fibers arising from the trochlear nucleus differ from those of any other cranial nerve in that the great majority decussate with those of the opposite side. After leaving the nucleus the fibers curve dorsally around the central gray mass surrounding the aqueduct to reach the anterior medullary velum in which the decussation occurs. They emerge from the dorsal surface of the anterior medullary velum on one side of its frenulum and immediately behind the inferior colliculus.

THE ABDUCENT NUCLEUS furnishes fibers to the external rectus. It lies in the pons close to the median line and subjacent to the upper part of the floor of the 4th ventricle. Its fibers pass downwards and forwards through the pons to emerge without crossing at the latter's lower border. The fibers of the facial nerve loop around the abducent nucleus (ch. 66).

The nuclei of the three ocular nerves receive fibers from (a) the pyramidal tract of the opposite side, (b) the medial longitudinal fasciculus through which the three nuclei are connected with one another, with the vestibular nucleus, with the spinal cord and probably with the facial nucleus. It has been suggested that fibers from the oculomotor nucleus may enter the latter nucleus and be then conveyed in the facial to the orbicularis oculi and the corrugator supercilii, (c) tectobulbar tract which relays to the three nuclei, impulses entering the superior colliculus from the optic tract and the visual cortex.

The eyes are said to be in a *position of rest* or in their *primary position* when their direction is maintained simply by the tone of the ocular muscles, that is, when the gaze is straight ahead and far away and not directed to any particular point in space. The visual axes are then parallel. When the eyes view some definite object they are turned by the contraction of the ocular muscles

and converged so that the visual axes meet at the observed object and an image of the object falls upon a corresponding point on each fovea (p. 1178). The closer the object to the eye the greater the degree of convergence (p. 1156). This movement of the eyes for the acute observation of an object is called *fixation*. The point where the visual axes meet is called the *fixation point* and the lines joining the latter to the fovea, i.e., the visual axes, are sometimes called the *fixation lines*. The widest limits of vision in all directions within which eyes can fixate is called the *field of fixation*. When surveyed by means of the perimeter it is found to be nearly circular with a diameter of about 100°. Its boundaries therefore lie well within the limits of the binocular visual field (p. 1167).

The eyeball is rotated in its socket (formed by the fascia bulbi) by the ocular muscles around one or other of three *primary axes* which intersect one another at right angles near the center of the globe. One axis is vertical, around it lateral movements (adduction and abduction) take place, i.e., in the horizontal plane. Another runs from before backwards and coincides with the visual axis, movements in the frontal plane (torsion or wheel movements) take place around it. The third is transverse, it is the axis of rotation for upward and downward movements, i.e., movements in the sagittal plane. Though the movements of the eyeball are essentially and for practical purposes rotary in character, a very slight translatory movement may take place as a result of movements of the lids and variations in the width of the palpebral fissure, closure and opening of the lids causing a displacement backward and outward, and forward and inward, respectively. A slight displacement at right angles to the rotary movement also takes place during contractions of the ocular muscles, the eyeball therefore executing what has been described by Berlin as a screw movement. Also for this reason the center of rotation of the eyeball is not an absolutely fixed point but varies slightly. For general purposes, however, it may be taken as the point of intersection of the primary axes. This point is on the visual axis about 13 mm. from the anterior surface of the cornea.

In table 100 the actions of the individual ocular muscles are given, but no normal movement is carried out by one of these muscles alone. Thus, when the eye is abducted, the external rectus and the two obliques act in unison to turn the eye outwards. The depressor and elevator components in the actions of the respective oblique muscles cancel one another. Similarly, adduction is effected by contraction of the internal rectus

acting with the superior and inferior recti. Again, the depressor and elevating actions of the latter two muscles neutralize one another. In looking upwards the eye is elevated by the combined action of the superior rectus and the inferior oblique. In looking downwards the inferior rectus and the superior oblique act together, the subsidiary action

unison, both turning in the same direction—*conjugate deviation*—and reciprocal innervation is extended to include muscle groups in the two eyes, thus indicating their control from a single center. Thus, stimulation of the posterior part of the 2nd frontal convolution causes conjugate deviation of the eyes to the opposite side. This involves contraction of the abductors and inhibition of the adductors of one eye and converse actions

TABLE 100

MUSCLE	MOVEMENT	INNERVATION	DIPLOPIA DUE TO OCULAR PARALYSIS POSITION OF FALSE IMAGE IN RELATION TO TRUE WHEN RIGHT EYE AFFECTED (APPLICABLE TO LEFT EYE IF RIGHT BE CHANGED TO LEFT AND VICE VERSA)	
Rectus	Superior	<i>Elevation</i> (IO), adduction (Inf R, Int R) intortion (SO)	Oculomotor	Above, to left of and tilted away from true image (crossed diplopia)
	Inferior	<i>Depression</i> (SO), adduction (Int R, SO) extortion (SO)	Oculomotor	Below, to left and tilted towards true image (crossed diplopia)
	Internal	<i>Adduction</i>	Oculomotor	Level with, parallel to and on the left of true image (crossed diplopia)
	External	<i>Adduction</i>	Abducens	Level with, parallel to and on the right of true image
Oblique	Inferior	<i>Extortion</i> (IR), elevation (SR), abduction (Ext R, SO)	Oculomotor	Above, to right of and tilted away from true image
	Superior	<i>Intortion</i> (SR), depression (Inf R), abduction (Ext R, SO)	Trochlear	Below, to right of and tilted towards true image
Levator palp sup	Elevator of eyelid antagonizes the action of the palpebral part of the orbicularis oculi	Oculomotor		

of the inferior rectus in turning the eye inwards being offset by the opposite action of the superior oblique. This compound action of the ocular muscles makes for smooth and steady movement and rapid fixation of the eyeball. It will be seen from figure 76 16 that the obliques and the superior and inferior recti when contracting individually produce a rotary or wheel-like movement, outward (extortion) or inward (intortion). When acting in pairs the rotary actions being in opposite directions antagonize one another so that normally no such movement occurs. It has been stated by Hering that every central influence governing eye movements, excitatory or inhibitory, reaches both eyes equally (Hering's law) causing contraction or relaxation of associated muscle groups.

The actions of the eye muscles follow the principle of reciprocal innervation. Thus, when the eye is turned outwards the external rectus and the two obliques contract while their antagonists (inferior, external and superior recti) are inhibited. The two eyes act in

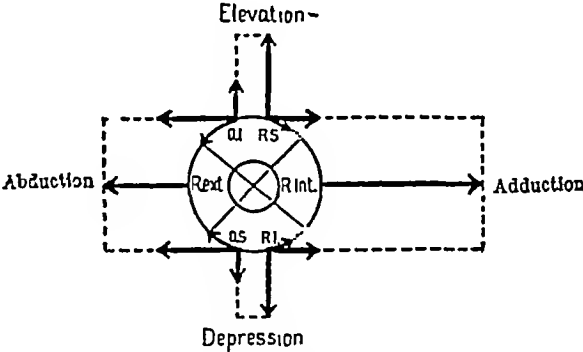


FIG 76 16 Diagram of right eye from the front to illustrate the actions of the ocular muscles O I, inferior oblique, RS, superior rectus, R ext, external rectus, R int, internal rectus, O S, superior oblique, RI, inferior rectus (from Fuchs after Marquez)

in the opposite eye (i.e. inhibition of the abductors accompanied by contraction of the adductors). Destruction of the cortical area results in the loss of the conjugate movement without paralysis of the individual muscles. The act of *convergence*, in which both eyes are adducted, is due to the conjoint contraction of the

internal recti. The center for this movement is probably in the central nucleus of the oculomotor nerve. A higher center for the movement is also situated in the frontal cortex.

The effects upon the eye movements of paralysis or weakness (paresis) of the ocular muscles. (a) *Limitation of movement* of the eye in the direction of the normal action of the affected muscle. (b) *Paralytic strabismus or squint.* When an effort is made to turn the eyes in the direction of the paralyzed muscle, the affected eye remains stationary or makes a smaller movement than does the sound eye. That is, it deviates in relation to the latter in a direction opposite to that of the normal action of the paralyzed muscle. The visual axes, therefore, do not bear their normal relationship to one another. This is called the *primary deviation*. If a screen is placed in front of the sound eye while an attempt is made to fixate the affected eye upon an object situated towards the side of the paralyzed muscle, the sound eye deviates in the direction of action of the latter. This *secondary deviation*, as it is termed, is greater than the primary deviation of the paralyzed eye. The greater deviation of the sound eye is attributed to the unusual effort exerted in the attempt to fixate the paralyzed eye, an unnecessarily strong motor discharge being transmitted simultaneously to the muscle of the sound eye which normally acts conjointly (conjugate deviation) with the paralyzed muscle.

(c) *Diplopia, false projection of the visual field.* If, as a result of weakness or paralysis of the muscles of one eye, or of an imbalance from whatever cause between the actions of the ocular muscles of the two eyes, the images do not fall upon corresponding retinal points, *diplopia* or *double vision* results. The image seen by the sound eye is called the *true image*, that seen by the affected eye is called the *false image*. The false image lies to one side, above or below the true image, depending upon the ocular muscle which is paralyzed. In the case of the oblique muscles and the superior and inferior recti, the false image lies above or below the true image—a little to one or other side and tilted towards or away from it (see table 100, p. 1183). The false image is always displaced in the direction of the normal action of the paralyzed muscle. Thus, in paralysis of the right external rectus the right eye is not turned outwards when the subject attempts to look at an object towards his right side. The image of the object falls upon the temporal half of the left macula and is therefore projected into the nasal half of the visual field of that eye. But, in the affected eye the image falls upon the nasal half of the retina and is therefore projected into the temporal half of the right visual field. The image seen by the right eye (false image) therefore lies to the right

of that seen by the left (true image). When the false image is on the same side of the true image as the affected eye the diplopia is said to be *simple* or *uncrossed*, if it lies on the opposite side of the true image the diplopia is said to be *crossed*.

If in a case of diplopia the sound eye is covered and the patient asked to turn his eyes, and to touch quickly an object placed to one side of his line of vision, but in the direction of the paralyzed muscle, he places his hand some distance from the object's true position. Normally, information concerning the positions of objects in space is to a large extent dependent upon proprioceptor impulses arising in the ocular muscles. When, for example, we look at an object straight in front of us an image falls upon the macula of each retina. When we look at an object to one side the eyes are turned so that the images fall again upon precisely the same areas—the maculae. The actual position of the object—whether in front or to one side—is made known to us by afferent impulses set up in the muscles as they turn the eyes into position. Such impulses serve also as a basis for the nice correlation between visual sensations and various body movements. When, as a result of paralysis of certain muscles, the eye does not move with the sound eye, the impulses arising in the muscles of the latter convey the impression, nevertheless, that such a movement has taken place. Let us say the outward movement of the right eye is paralyzed and the left, sound eye, is covered, when the latter turns to the right the patient believes the paralyzed eye does so to the same degree. The image in this eye, which continues to look forward, falls upon the nasal side of the retina. Since earliest experiences have taught him that an image falling upon the nasal part of the retina when this eye is rotated outwards represents an object well over to the right, he falsely projects the image into this position, and makes the appropriate movement of the hand in an attempt to touch the object.

If strabismus is congenital or of long standing diplopia is not, as a rule, experienced. This is because though the images do not fall on anatomically symmetrical corresponding points in the two eyes, an area is developed in the peripheral retina of the squinting eye which assumes the function of a fovea. This pseudofovea or false macula, as it has been called, corresponds physiologically to the fovea of the sound eye. Fusion of the two images occurs and stereoscopic vision suffers little if at all. When squint of long standing is corrected surgically to bring the visual axes parallel, diplopia results since the true fovea of the corrected eye does not correspond functionally with the fovea of the normal eye.

THE EAR ANATOMICAL OUTLINE SOUND GENERAL PRINCIPLES

ANATOMICAL OUTLINE

In order that the reader may follow with the least effort the account of the physiological mechanisms of hearing, the anatomy of the auditory apparatus and especially those features having a direct bearing upon function will be briefly described

The *external ear* comprises the cutaneous and cartilaginous appendage known as the *auricle* or *pinna*, and the short passage—the *external auditory meatus* leading into and penetrating the temporal bone (fig 77 1)

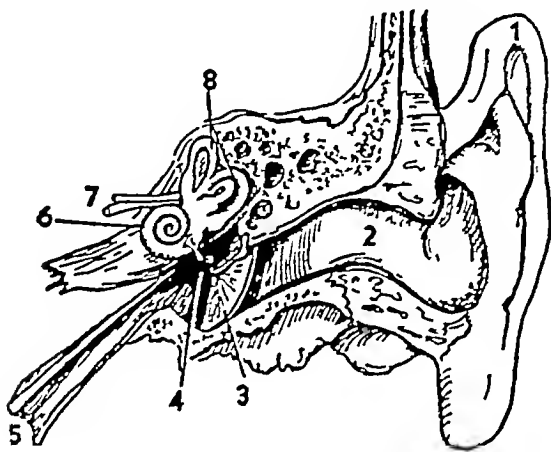


FIG 77 1 Plan of the ear (Redrawn and modified from Arnold) 1, the auricle (or pinna), 2, external auditory canal (or meatus), 3, the tympanic membrane (sectioned), 4, the tympanic cavity (middle ear), the chain of ossicles lies just above the pointer, 5, Eustachian tube, 6, cochlea, 7, acoustic nerve, showing cochlear and vestibular divisions, 8, semicircular canals

The external auditory meatus is an S shaped canal, being directed inward, upward and forward, then, upward and backward and finally, forward and slightly downward to end blindly at the drum membrane. The inner 16 mm. or so of the canal (pars ossea) forms a tunnel in the temporal bone. The wall of the outer 8 millimeters of the meatus is composed of cartilage (pars cartilaginea), the entire canal, including its inner blind end formed by the drum membrane, is lined by skin.

The *middle ear*, *tympanic cavity* or *drum* is a chamber measuring about 15 x 5 x 2 mm and situated within the petrous bone. It is lined by mucous membrane. Except over the tympanic membrane and part of the inner wall, the epithelium is ciliated. The outer wall of the tympanic cavity is formed mainly by the flexible *tympanic (drum) membrane*, while its inner wall which separates it from the internal ear is composed of bone,

except for the *oval* and *round windows* (*fenestra vestibuli* and *fenestra cochleae*). The inner wall presents a rounded eminence—the *promontory*—caused by the projection laterally of the basal turn of the cochlea. The oval window lies above the posterior part of the promontory. The round window which is closed by a delicate membrane is situated at the bottom of a deep hollow or niche lying just below this end of the promontory. The oval window lodges the footplate of the stapes. The tympanic cavity is filled with air which is maintained at atmospheric pressure by means of the *Eustachian tube (auditory tube)* which runs from the lower part of its anterior wall to the nasopharynx. The lower opening of this tube is closed at ordinary times but is dilated during swallowing by contraction of the *salpingopharyngeus* and *dilatator tubae* muscles.

The tympanic membrane is not quite circular, measuring about 10 mm in height and 9 mm in width. It is placed obliquely, being directed from above downward and forwards to form an angle of 55° with the anterior and inferior walls of the external auditory meatus. It possesses three layers, the outermost being of skin and the innermost of non-ciliated mucous membrane. The middle layer or *membrana propria* is composed of two sets of fibers—a radial and a circular—arranged somewhat like the threads of a spider's web. The drum membrane is drawn inwards at the center and along the attachment of the handle of the malleus, its outer surface is therefore concave, its inner surface convex. The point of greatest concavity, which corresponds with the tip of the handle of the malleus, is called the *umbo* (fig 77 2).

The drum membrane is observed during life by means of a light thrown into the meatus through an aural speculum. The healthy membrane viewed in this way is pearl gray in color, pinkish or faintly yellow. Above and anteriorly near its circumference the membrane presents a small white spot caused by the projection of the lateral process of the malleus, from this point a faint ridge corresponding with the handle of the malleus extends to the umbo. Two faint folds—the *anterior* and *posterior malleolar folds*—extend forward and backward, respectively, from the lateral process of the malleus, enclosing a small triangular area. The part of the tympanic membrane within this area is thin and lax, it is known as *Shrapnell's membrane* or the *pars flaccida*. The rest of the drum membrane is tight and glistening (*pars tensa*). The lustre of the membrane gives rise to a bright triangular area, the "cone of light." This is situated with its apex at the umbo and its base directed downwards and forwards. For convenience in describing the position of a lesion of the

drum membrane it is mapped out into quadrants by a line represented by the handle of the malleus and its continuation downwards and by another passing through the umbo at right angles to the first.

Three articulated miniature bones—the auditory ossicles—stretch across the tympanic cavity from the drum membrane to the oval window (fig 77.3) These are named, somewhat imaginatively from their shapes, the malleus (hammer), the incus (anvil) and the stapes (stirrup). The *malleus* (8–9 mm long) consists of a handle (or manubrium) which is attached along the upper half of the vertical diameter of the tympanic membrane, a head (or capitellum) and two processes, a lateral and an anterior. The *incus* is shaped like a premolar tooth, the anterior surface of its body presents a saddle shaped facet which articulates with the

The tympanum contains two minute muscles—the *tensor tympani* and the *stapedius*. The former arises from the roof of the cartilaginous part of the Eustachian tube and from the adjacent part of the great wing of the sphenoid bone, its tendon turns laterally to be inserted into the medial edge and anterior surface of the handle of the malleus (fig 77.3). When it contracts it pulls the handle of the malleus inwards, thereby preventing excessive displacement outwards of the tympanic membrane. The *stapedius* takes its origin from an eminence (eminencia pyramidalis) on the posterior tympanic wall and passing forwards is inserted into the posterior surface of the neck of the stapes, its action pulls the head of the ossicle backwards, thus tilting the anterior edge of the base outwards, i.e., towards the tympanic cavity, and reducing the pressure upon the

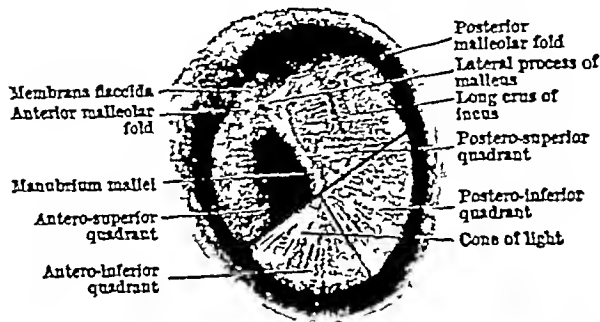


FIG 77.2 Left tympanic membrane (as viewed from the external acoustic meatus) (From Cunningham's Anatomy.)

posterior aspect of the head of the malleus. The lower part of the articular facet is hollowed out for the reception of the prominent inferior margin of the facet on the posterior surface of the head of the malleus—the so-called *spur* or *cog tooth* of the malleus. The incus has a long and a short process, the former projects downwards and, turning inwards at the tip, connects through a ball and socket joint with the head of the stapes. The stapes shows a base (footplate of the stirrup) and two curved crura which join laterally to form an arch, a rounded eminence called the head arises from the center of the latter. The slight constriction between the head and the junction of the crura is called the neck. The head of the stapes, as just mentioned, articulates with the incus. The base fits into the oval window and is coated with cartilage which is connected around its circumference to the margins of the oval window by a ring of elastic fibers, known as the *annular ligament*.

The auditory ossicles are connected to the walls of the tympanum by five ligaments. Three of these are attached to the malleus—the *anterior*, *superior* and *lateral ligaments of the malleus*, a fourth connects the short process of the incus to the posterior wall of the tympanum, the fifth is the annular ligament of the stapes just mentioned.

perilymph. These muscles are under reflex control, contracting to sound stimuli over the entire range of audible frequencies. The thresholds for these reflexes are lowest for tones of from 2000 to 4000 c.p.s. (rabbit). A sound in one ear causes increased tone of the tensor tympani of the opposite ear. This muscle also contracts during yawning. The tensor tympani is supplied by the fifth nerve, the stapedius by the seventh.

Two main views have been expressed concerning the functions of the intra-aural muscles. Some believe that they are *protective* in action, their contractions tending to reduce the amplitude of the vibrations of the tympanic membrane and ossicles particularly to low tones, thus protecting the delicate structures of the internal ear from injury. The other view, generally referred to as the *accommodative theory*, holds that the stapedius and tensor tympani act to "tune up" the transmitting mechanism of the middle ear and thus to increase its sensitivity to any given vibration frequency. It has been shown, however, by H. Wiggers in the guinea pig that the transmissibility of low tones (below 1000 cycles) to the internal ear is *reduced* by contraction of the stapedius and tensor tympani, the transmission of medium tones (1300 to 1800 cycles) is slightly enhanced while that of high tones (over 2000 cycles) is

unaffected In the human subject the transmission of tones within the range of a conversational voice and of all other tones of low pitch is reduced These observations indicate that the function of the intra-aural muscles is mainly protective, any direct effect upon auditory acuity is limited to a small pitch range However, their action in reducing the efficiency of the transmission system for low tones may indirectly, by diminishing the masking effect (p 1196) of such tones, increase the acuity for higher frequencies

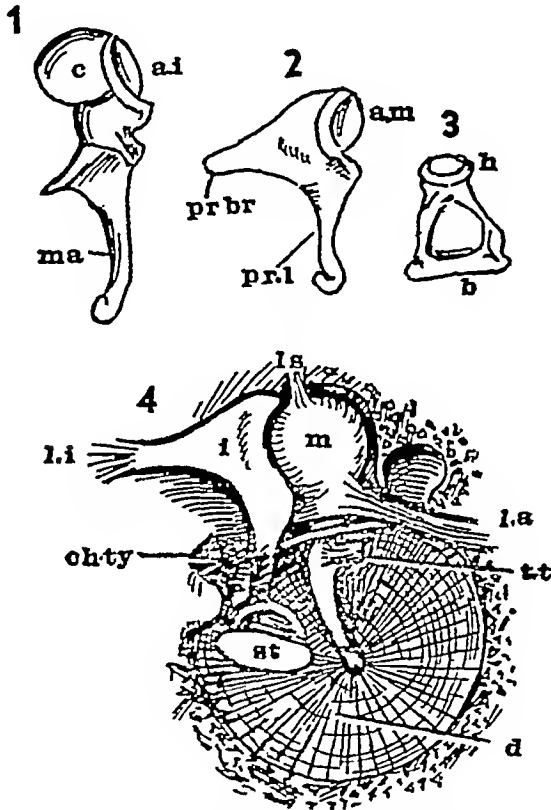


FIG 773 The auditory ossicles 1, left malleus viewed from outer side, c, head, a1, articular surface for incus, ma, handle. 2, left incus, pr br, short process, pr l, long process which articulates with stapes a m articular surface for malleus 3, left stapes, h, head, b, base or "footplate", 4, the middle ear viewed from the inner aspect and showing ossicles in position, d, drum membrane, i, incus, m, malleus, st, stapes, ch ty, chorda tympani nerve, ls, ligament of malleus, li, ligament of incus, tt, tendon of tensor tympani muscle

The internal ear or labyrinth, situated within the temporal bone on the inner side of the middle ear, contains the auditory sense organs The latter lie within a spiral canal called the cochlea (L a snail's shell) The canal makes $2\frac{1}{2}$ turns round a central pillar of bone called the modiolus The smallest turns of the cochlea are at its apex or cupula which is directed forwards and laterally, the largest turns at the base which looks backwards and inwards and forms part of the outer wall of the internal auditory meatus A ledge of bone

(lamina spiralis ossea) winding around the modiolus like the thread of a screw-nail divides the spiral canal incompletely into two parts The partition is completed by a membranous structure—the basilar membrane—which extends from the tip of the lamina spiralis ossea to the outer wall of the canal (see fig 774) A second membrane—Reissner's membrane—stretches from the upper surface of the bony lamina to a point a short distance above the outer attachment of the basilar membrane The original osseous canal is divided in this way into three spiral compartments or galleries The gallery below the basilar membrane is called the

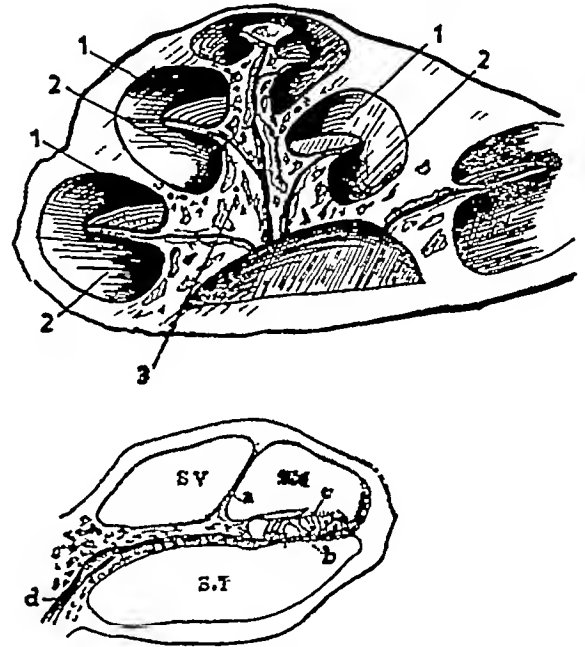


FIG 774 Upper, a view of the osseous cochlea divided through the middle 1, scala vestibuli, 2, scala tympani, 3, modiolus, showing tunnels transmitting branches of auditory nerve Lower, enlarged sketch of one turn of the cochlea (redrawn after Quain) S V, scala vestibuli, S M, scala media (cochlear duct), S T, scala tympani, a, Reissner's membrane, b, basilar membrane, c, organ of Corti, d, auditory nerve

scala tympani, the one above Reissner's membrane, the scala vestibuli, while the one enclosed between the two membranes is known as the scala media, the cochlear duct or the membranous cochlea The membranous cochlea is filled with a fluid known as endolymph, while the osseous canals, i.e., the scala vestibuli and scala tympani, are filled with perilymph The scala vestibuli and scala tympani communicate with one another at the cupula of the cochlea through a small aperture called the helicotrema The scala media ends here as a closed sac, but its basal extremity communicates with the non-auditory labyrinth through the canalis reuniens or duct of Hensen (p 975) The perilymph and endolymph have different origins The former is actually cerebrospinal fluid, the osseous canals communicating

with the subarachnoid space through the *ductus perilymphaticus* which passes from the floor of the vestibule to the posterior fossa of the skull. The origin of the endolymph is unknown, it is possibly a secretion or transudate furnished by the stria vascularis. The scala vestibuli opens out near the central part of the labyrinth into an oval osseous chamber ($6 \times 4 \times 4$ mm) called the *vestibule*. This chamber contains the utricle and saccule, the osseous semicircular canals open into its posterior part (p. 975). Its lateral wall separates it from the middle ear and contains the oval window.

The auditory receptors. The sensory cell, together with sustentacular elements constitute a structure known as the *spiral organ of Corti* (fig. 77.5). This lies within the scala media, occupying the inner half or so of the basilar membrane. It presents towards its inner part two rows of elongated epithelial elements of

separated from the basilar membrane by cells arranged in several rows—the *cells of Deiters*—which send slender processes between the rows of the sensitive cells. That part of the organ of Corti lying on the outer side of Deiters' cells and the outermost row of hair cells is composed of several layers of columnar cells—the *supporting cells of Hensen*.

The *tectorial membrane* is a delicate, almost homogeneous structure, somewhat paddle-shaped in transverse sections of the scala media, it overlies the inner half or more of the spiral organ. Arising from a point near the base of the lamina spiralis ossea it ends laterally in an irregularly fringed or scalloped free margin. The hairs of the sensory cells of all four rows are embedded in its under surface.

The structure of the basilar membrane. The inner part (*zona arcuata*) of the basilar membrane where it

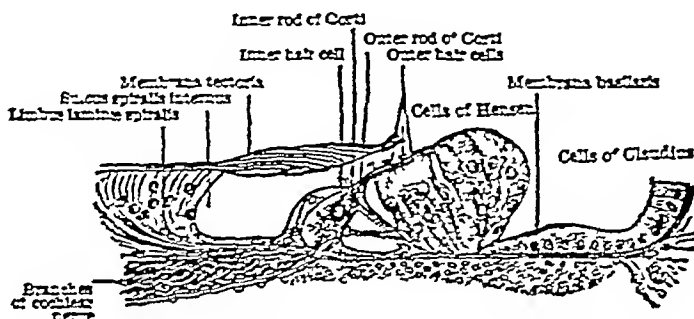


FIG. 77.5 Section across the organ of Corti (After Retzius)

peculiar shape—the *inner and outer rods* (or *pillars*) of *Corti*—which stand with their expanded bases planted upon the basilar membrane. The bases of the rods of the two rows are separated by an interval, but above the membrane their bodies incline towards one another and meet at an angle, these extremities fit into one another in a manner much as one clasps a fist with the other hand. They form by this arrangement a triangular tunnel whose floor is covered by two nucleated scraps of protoplasm which actually are the undifferentiated parts of the cells from which the corresponding rods have developed. The arches of *Corti* increase progressively in height from the basal to the apical turns of the cochlea.

The *hair cells* are the essential sensory elements. In section they are seen as four rows of short columnar cells, one row lying on the inner side and three on the outer side of the corresponding rod. The *inner hair cells* number about 3500, their free ends are in line with the surfaces of the rods, but their bases extend only half way to the basilar membrane. The free surface of each hair cell of both inner and outer rows is surmounted by some twenty hair-like processes. One or two rows of cells support the bases and inner aspects of the inner hair cells. The *outer hair cells* are longer than the inner and more numerous (12,000), they are

supports the spiral organ is thin, its outer thicker part (*zona pectinata*) is covered by a single layer of columnar cells named the *cells of Claudius*. The substantia propria of both zones of the membrane show numerous fibers—the *auditory strings*—embedded in a homogeneous ground substance. The strings in the outer zone are straight and smooth and run for the most part transversely, those of the inner zone are thinner and arranged in the form of a net. The total number of fibers in the human cochlea is 24,000, according to the generally accepted estimate of Retzius. The basilar membrane when uncoiled shows a gradual taper; it measures about 30 mm. in length in man. It is about three times wider at its apical than at its basal end. The auditory strings vary in length in a corresponding manner, being from 65μ to 160μ long at the base and from 350μ to 500μ at the cupula. The outer end of the basilar membrane is attached to the wall of the cochlear canal throughout its entire length by a ligamentous band called the *external spiral ligament*. This appears in sections of the canal as a conical structure composed of radiating fibers, it is relatively thick and strong in the basal turns but becomes progressively narrower in the direction of the cupula where it dwindles to a delicate strand.

The external spiral ligament is covered by cubical

epithelium Situated between the epithelial cells and the underlying fibrous tissue are numerous small blood vessels and capillary loops which together form a structure named the *stria vascularis*. The tissue surrounding the vessels contains pigment of varying amount depending upon the species. The *stria vascularis* is thought by some to have a secretory function.

Innervation of the hair cells The cochlear division of the auditory (acoustic) nerve emerges as a number of fine filaments from the base of the modiolus in the internal auditory meatus. These are the central processes of the bipolar cells of the *spiral ganglion of the cochlea* which lies in the *spiral canal of the modiolus*, the latter twists through the bone along a line nearly corresponding with the origin of the *lamina spiralis ossea*. The peripheral processes (axons) of the bipolar cells proceed outwards in fine canals in the spiral lamina to its outer edge. The nerve filaments terminate as fine arborizations, some around the bases of the inner hair cells, others after crossing Corti's tunnel around the bases of the outer hair cells.

The non-auditory part of the labyrinth is described in chapter 65.

THE AUDITORY PATHWAY The cochlear division of the 8th nerve after issuing from the internal auditory meatus, crosses the posterior fossa of the skull in close association with the vestibular division (p. 975) and the facial nerve. Reaching the lower border of the pons the cochlear fibers divide into two groups. The fibers of one group end around cells in the *ventral cochlear nucleus* situated upon the ventro-lateral aspect of the *restiform body* (fig. 77.6), those of the other group around cells of the *dorsal nucleus (tuberculum acusticum)* lying upon the dorso-lateral aspect of the *restiform body*. Secondary auditory neurons are situated in both these nuclei. The fibers issuing from the ventral nucleus pass medially, forming the *trapezoid body* and, crossing with those of the opposite side, ascend in the *lateral lemniscus*. (Some fibers from the ventral nucleus give off collaterals to the nucleus of the trapezoid body and to the superior olive, from these nuclei tertiary neurons convey impulses through the *medial longitudinal fasciculus* to the nuclei of the *oculomotor*, *trochlear*, *abducens* and *spinal accessory nerves*.) Fibers arising from the cells of the *dorsal nucleus* pass around the dorsal aspect of the *restiform body*, and proceed medially in the floor of the 4th ventricle, where they appear as well defined white strands—the *striae acusticae* (or *striae medullares*). Reaching the midline they cross to the opposite side, and ascend with those of the ventral nucleus in the *lateral lemniscus*. Some fibers of the dorsal nucleus end in the nucleus of the superior olive, and some from both the dorsal and ventral nuclei join the *lateral lemniscus* of the same side. The *lateral lemniscus* as just indicated is composed of the secondary neurons of the auditory pathway and ascends in the *reticular formation* of the pons, its fibers ending in (a) the *substantia nigra*, (b)

the *inferior colliculus* and (c) the *medial geniculate body*. Just as the *superior colliculus* serves as a center for visual reflexes so the *inferior colliculus* is a center for auditory reflexes. It is not concerned with auditory sensations. It is connected by descending tracts with the nuclei of the brain stem and spinal centers (*tecto-spinal tract*). The *medial geniculate body* is the sub-cortical or primary auditory center.

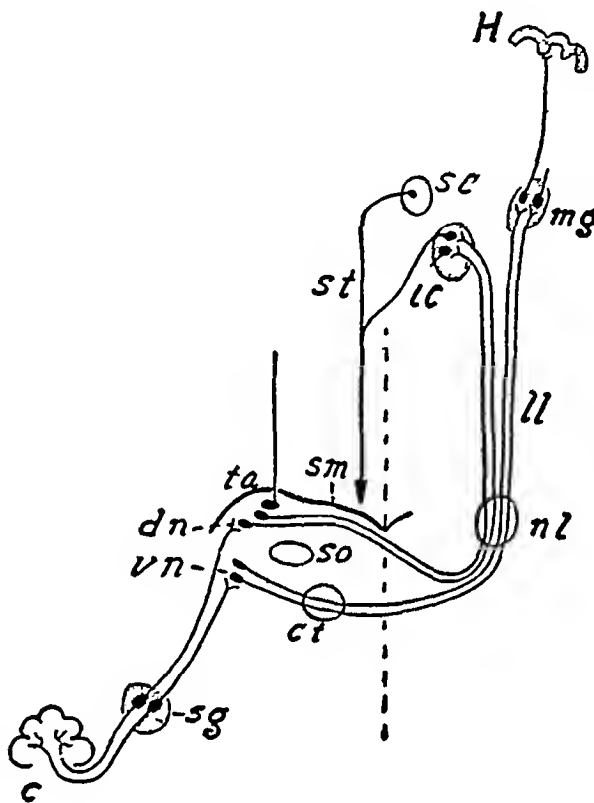


FIG. 77.6 Diagram of the auditory pathway. H, Heschl's gyrus, sc, superior colliculus, ic, inferior colliculus, mg, medial geniculate body, st, tectospinal tract, sm, striae medullares, ta, tuberculum acusticum, dn, dorsal cochlear nucleus, vn, ventral cochlear nucleus, so, superior olive, ct, corpus trapezodeum, sg, spiral ganglion, ll, lateral lemniscus, nl, nucleus of the lateral lemniscus, c, cochlea. Collateral connections with so, ct and nl are not shown. Mid-line, dotted line.

The auditory radiation The tertiary auditory neurons arise in the *medial geniculate body*,¹ ascend in the posterior limb of the *internal capsule* external to the fibers of the *optic radiation*, and end in the cortex of the *superior temporal convolution* (*Heschl's gyrus*.)

The peripheral organ shows a point to point localization in the cochlear nuclei and temporal cortex analo-

¹ There is some evidence that the secondary neurons make connections with cells in the superior olivary nucleus or in the nucleus of the lateral lemniscus. These would therefore be tertiary neurons, while those arising in the *medial geniculate body* would be of the fourth order.

gous to that described for the retina. Those fibers from the apical turn are distributed mainly to the ventral parts of the cochlear nuclei while those from the basal turn terminate in the dorsal part of the nuclei, the endings of fibers from the middle coil occupy an intermediate position. A certain localization of fibers from the coils of the cochlea have been determined in the medial geniculate body, and projection from this structure to the cortex occurs in orderly fashion. Fibers from the medial part of the subcortical center pass to the depth of the Sylvian fissure, those from the lateral part to the region nearer the lip of the fissure (see ch 68 and figs 68.3 and 68.7)

SOUND, GENERAL PRINCIPLES

All sound arises as a series of vibrations. The sound is transmitted through an elastic medium (e.g., air) as a train of alternating variations in pressure. For example, when the prongs of a tuning-fork are struck they vibrate, moving rapidly to and fro with a pendular motion and producing

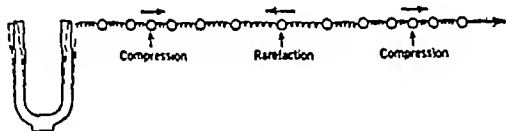


FIG 77.7 Showing how one prong of a tuning fork compresses and rarefies the air. The circles represent little portions of air, while the springs are intended to show the elasticity of air. (After Watson, *Sound*)

alternate compressions and decompressions (rarefactions) of the air in contact with their flat surfaces. The train of pressure alterations thus set up are transmitted through the air in all directions, reaching the drum membrane, they force it into vibration (fig 77.7). The pendular movement of the tuning fork is termed a *simple harmonic motion*. This movement is imparted to the air particles in contact with the prongs of the fork and through these to contiguous particles and so on through the medium. Simple harmonic motion may be illustrated by a mechanical model. In figure 77.7 the particles of air (O) are represented as connected by springs. When the prong of the fork moves outwards the first spring is compressed and a movement thereby transmitted to the attached particle. The movement of the particle compresses the second spring and through it the corresponding particle, and so on through the series. The springs are not compressed all at the same instant but in succession, the movement of each particle occurring a little time after that of the particle immediately preceding. When the prong swings in

the opposite direction the springs are stretched, the particles moving in reverse order, the movement showing a similar lag from particle to particle. Thus, the particles oscillate in a regular to and fro manner. Simple harmonic motion is illustrated by the *projection of a circular movement upon the diameter of the circle*. For example, if while a body is moving with constant velocity in a circle one observes it from a distance with the eye in the plane of the circle, it appears to oscillate back and forth like a shuttle along a straight line, that is, along the diameter of the circle. In order to indicate the location of the oscillating particle on its linear path at any instant, another particle, purely imaginary, is supposed to be traveling in a circle and always keeping vertically above the actual one. The orbit of this imaginary particle is called the *circle of reference* (see fig 77.8). The distance which the particle travels from its equilibrium point (which is the center of its linear

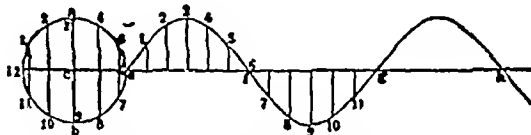


FIG 77.8 Illustrating simple harmonic motion and circle of reference. (After Wilkinson and Gray)

path), i.e., half a full excursion, will, of course, be proportional to the radius of the circle and gives the *amplitude* of the vibration. The position of the particle in the circle of reference is called the *phase* of its motion. A complete movement, i.e., from one phase to the next corresponding phase is termed a *cycle* or *double vibration* (d.v.) and the number of cycles per second (c.p.s.) is the *vibration frequency*. The distance from one particle to the next one in the same phase gives the wave length. The compression or stretching of a spring in the model corresponds to the movement of a particle of the medium in the same or in the opposite direction, respectively, to that in which the sound wave is moving, in their movement with the wave the air particles are closer together (compression), during the opposite movement they are further apart (rarefaction). Simple harmonic motion is the commonest type of motion set up in air particles by vibrating bodies and *all musical tones are due to such motions or to a combination of them in different phases* (p. 1194).

Though, as stated above, sound is conducted as a series of compressions and expansions of the air, it is

customary to represent it graphically as travelling like a series of water waves. But it must be pointed out that this is merely a useful convention derived from the fact that if we attach a writing point to a pendulum or tuning-fork executing simple harmonic motion, and have it inscribe its excursions upon a moving surface, a series of such waves will be drawn. Such a simple regular type of wave produced by a tuning-fork and shown in figure 77 8 is known as a sine wave, because when different points upon the wave are correlated with the positions of the moving particle on the circumference of the circle of reference, the ordinate at each point on the wave is proportional to the sine of the angle which the radius joining the particle on the circumference makes with the diameter of the circle. Simple or pure tones are all produced by waves of this type.

The *velocity of sound* varies with the elasticity of the medium and inversely with its density. Increased elasticity raises the velocity (the imaginary springs between the particles being stiffer respond more rapidly), greater density slows the rate of transmission owing to the greater inertia of the particles. The velocity of sound *v* aches is about 1100 feet per second in air, 4,700 per second in water, 13,000 per second in wood and 16,500 in steel. The velocity of sound in a medium is obtained from the equation $v = \sqrt{E/d}$, in which E = elasticity and d = density. The wave length of a given tone is readily obtained by dividing its frequency of vibration (number of cycles per second) into the figure expressing its velocity. Thus the wave length of a sound transmitted through air and with a frequency of 550 cycles per second is $(1100/550 =) 2$ feet.

Sound waves undergo *absorption, reflection, refraction and diffraction* in a manner closely similar to that exhibited by light rays (chapter 75). For example, sound waves travelling in one medium upon striking another possessing a different density or elasticity are in part absorbed, in part reflected and in part transmitted, the proportions disposed of in each of these ways depending upon the differences between the properties of the two media. Sound waves falling upon a substance such as felt, absorbent cotton or porous fiber board are largely absorbed, whereas those striking water or a hard smooth surface such as glass, steel or a plastered wall are nearly all reflected. A suitably curved concave surface of some hard material is capable of converging the sound waves to form a sound "image." Echoes and reverberations are caused by sounds thrown back towards their source from reflecting surfaces, and whispering galleries in many instances owe their properties to the architecture of the hall—the walls or ceiling acting as a convex "mirror" to concentrate the sound waves within a small area. Sound is diffracted to a much greater extent than is light, owing to the much greater length (1,000,000 times at least) of the waves of the former. For example, sounds coming through a

small open window spread out and fill the whole room just as though the aperture were itself a source of sound.

RESONANCE, FREE AND FORCED VIBRATIONS

Some sounding bodies, such as the strings of a harp, vibrate freely at a certain frequency after the force which set them in motion ceases to act. They are said to be tuned to a certain pitch or frequency, this is called their *natural frequency*. When sound waves corresponding to the natural frequency of such a body fall upon it, it vibrates and gives out the same note as it would if it were actually struck. This phenomenon is known as *resonance* or *sympathetic vibration*. For example, should one raise the dampers of a piano by depressing the loud pedal and sing a note near by or sound one upon a violin or other instrument, that string of the piano vibrates which if actually struck would give the same note—it "sings" in unison with the original tone. Though a resonator gives the maximum response to a sound having the same frequency it also vibrates less strongly to other frequencies a little higher or lower than its own.

A body which possesses no natural frequency is called *aperiodic*. When such is made to vibrate or when one which possesses resonating properties is made to vibrate at a frequency other than its natural frequency, as when a tuning-fork is driven by an alternating electric current, the vibrations are said to be *forced*. In setting up forced vibrations a much greater amount of energy must be expended than in the production of sympathetic vibrations. A vibrating tuning-fork, for example, while it will readily set up vibrations in another fork of its own frequency some little distance away must, in order to cause a body such as a table top to vibrate, be brought into direct contact with it. Vibrations forced in this way stop when the motion which set them up ceases, this phenomenon is due to *damping*.

THE CHARACTERISTICS OF SOUND

Sounds may differ from one another in three particulars, namely, intensity, pitch and quality or timbre.

INTENSITY Sound intensity and loudness are not synonymous terms. The former is a purely physical value, whereas the latter refers to the auditory sensation. The sensitivity of the ear varies with pitch, therefore of two sounds having the same intensity but of different pitches, one

may be much louder than the other. Intensity refers to the energy (or power) of the sound waves and is proportional to the square of the pressure variations—the amplitude—of the waves, it may be quite independent of the auditory sensation. For example, a tone of a certain high pitch while inaudible, however intense, to the human ear, may be audible to some animals. Sound intensity is therefore given in physical units, namely, the number of ergs per second (or microwatts) passing through an area of 1 sq cm.² Or it may be expressed as pressure, namely, as dynes per sq cm. An absolute unit is less generally useful, however, than one which indicates *differences* of intensity. It is

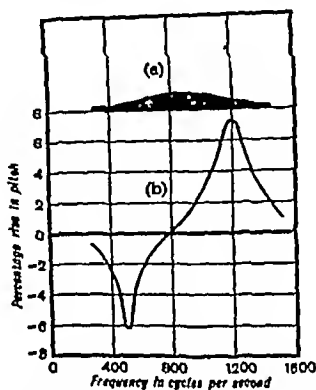


FIG. 77.9 Illustrating the effect of fatigue upon pitch. (a) The blackened area indicates the distribution of fatigue among nerve endings due to a 800 ~ tone. Hence the sensation due to a 1200 ~ tone is modified as shown by the curve (b), namely, sharpening of pitch for tones of higher frequency and flattening for lower tones. (From Beatty after Békésy.)

now customary to use a scale in which the intensity of a sound is expressed as the logarithm to the base 10 of the ratio of two intensities, this unit is called the *bel* (after Graham Bell, the inventor of the telephone). The number of bels by which one sound intensity exceeds the other is called the *intensity level*. For example, if the intensity of the sound is increased ten times, say from 10^{-10} to 10^{-9} microwatts per sq cm, the intensity level is raised by 1 bel, a hundred-fold increase in intensity corresponds to a rise in the intensity level by 2 bels, of a thousand-fold by 3 bels and so on. The reference intensity, i.e., the standard with which a given intensity is compared, may be anything we

² 1 dyne = force which, acting on a mass of 1 gram for 1 second, produces an acceleration of 1 cm per sec.
1 erg = work done by 1 dyne acting through 1 cm.
1 watt = 10^7 ergs per sec., a microwatt = 10^{-6} watt.

choose, since the bel is a purely relative value. The otologist or physiologist takes as his reference, the intensity of the faintest audible sound in a perfectly quiet room. The sound at reference intensity must be of the same pitch as that of the sound with which it is compared. The following intensity values of a number of common sounds are given by Beatty. The sound of leaves rustling in a gentle breeze has an intensity 10 times that of the faintest audible sound, a whisper at a distance of four feet 100 times, sounds in a quiet street 1000 times, a conversational voice at twelve feet 100,000 times, a loud peal of thunder 10,000,000 times, and sounds at the limit which the ear can endure, 10,000,000,000,000 times. The corresponding intensity levels are 1, 2, 3, 5, 7 and 13 bels, respectively. For physiological or clinical work, the bel is inconveniently large. A unit of $\frac{1}{10}$ bel, called the *decibel* (abbrev. db) is used.³

Pitch is that property of sound which enables one to place a tone at a definite level in the musical scale. It is dependent mainly, though not entirely, upon the vibration frequency, i.e., upon the number of cycles falling per second upon the ear. That this property is not entirely a question of vibration frequency is shown by the fact that two sounds of the same frequency but at different intensity levels may be judged by the ear as differing slightly in pitch.⁴ When the intensity level of a

³ The *sensation level* is defined as the number of decibels that a sound is above the audible threshold of a sound of the same pitch. If the hearing is normal the sensation levels and intensity levels coincide, but are different if the threshold of hearing is raised. The *loudness level*. The intensity level and sensation level and their unit the decibel can only be used for expressing differences between sounds of the same frequency for, as we have seen, loudness varies with pitch. The loudness level is defined as the intensity level of a sound of 1000 cps. At this frequency but at no other do the loudness level and the intensity level coincide. The loudness level of any other pitch is determined by sounding a 1000 cycle tone—the *reference tone*—and raising its intensity until the two sounds are of the same loudness. The number of decibels by which the reference tone exceeds its audible threshold when the two tones are matched gives the loudness level. The energy of a 1000 cycle tone at the threshold of normal hearing is 10^{-16} watts per sq cm, this is therefore the zero or reference intensity on the loudness scale. The unit of the loudness level is sometimes called the *phon*. When the sound whose loudness is being measured has a frequency of 1000 cps, decibel and phon are interchangeable terms, but at all other frequencies the latter term is specifically applicable. For example, when a tone has a value of 1 phon, its loudness is 1 decibel above the reference intensity of a 1000 cycle tone, not above that of a tone of its own frequency.

⁴ There is no relation between intensity and frequency. The vibrations of a tuning fork, for example,

tone is raised it tends to become "flat" though its frequency remains the same, the pitch may be lowered at very high levels of intensity by as much as half a tone. The alteration in pitch in such an instance is attributed to the greater degree of tension exerted upon the resonating fibers of the basilar membrane (p 1206), by the louder tone and the rise, in consequence, of their natural frequency of vibration. For example, if fibers which resonate to C are made to vibrate strongly to a loud sound which would ordinarily give that pitch, their frequency rises while the frequency of fibers which ordinarily respond to B is increased to a corresponding extent, they now respond to the loud C.

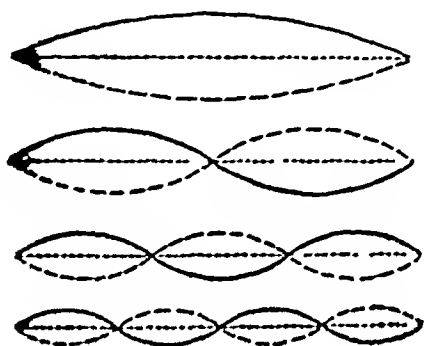


FIG 77 10 Shows the form assumed by a string vibrating as a whole and in two, three, and four equal parts, respectively

Impulses reaching the cerebral centers are interpreted as having been caused by a tone which under usual circumstances would set the B fibers into vibration (see Resonance Theory p 1205)

The pitch of a note of constant frequency may also change as a result of auditory fatigue (see fig 77 9). When the ear is fatigued by a tone, let us say of 800 c.p.s.,⁵ the end organs which are accustomed to respond to this frequency fail to do so or do so inadequately, while those on either side are little affected. As a result, the point of maximal stimulation (p 1206) undergoes an apparent shift in one or other direction. A tone of a frequency a little lower than 800 c.p.s. therefore is flattened, one a little higher in frequency is sharpened.

THE HARMONIC SERIES AND THE DIATONIC SCALE
A body which executes a simple harmonic or pendular motion (p 1190), such as the prong of a tuning-fork

become weaker and weaker, i.e., of smaller and smaller amplitude, after it is struck but their frequency remains unaltered.

⁵c.p.s., is the abbreviation for cycles per second. It has the same meaning as dv (double vibrations) per second.

or a stretched string, vibrating as a whole emits a pure or simple tone and will inscribe a sine curve. Simple tones whose frequencies are such as to form a series in which the higher frequencies are simple multiples of the lowest constitute an *harmonic series*. The ratios of the frequencies are represented by the numbers 1, 2, 3, 4, 5, 6, etc. For example, the frequencies of a series of tuning-forks of which the first vibrates at 250 cycles per second, will be 250, 500, 750, 1000 cycles per second, and so on. Or again, when a string or wire stretched tightly between two fixed points is struck so that it vibrates *as a whole*, it gives out the lowest tone of which it is capable. This is called its *first harmonic* or *fundamental tone*. If a bridge is placed exactly beneath its center, either half of the wire vibrates at just double the previous frequency, if divided into thirds the frequency is trebled. Divisions into quarters, fifths, sixths, etc., give corresponding frequencies. A tone having a frequency double that of the fundamental is called the *second harmonic* or *overtone* or the *octave*.

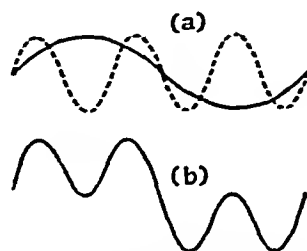


FIG 77 11 (a) Wave-form of pure tone of frequency f (solid curve) and of frequency $3f$ (dotted curve) (b) The compound wave

These considerations lead us to the important *law of lengths*, which states that the number of vibrations per second of a stretched string is inversely proportional to its length. The *laws of diameters*, of *densities* and of *tension* may be summed up in the statement that the vibration frequency is inversely proportional to the diameter of the string and to the square root of its density, and is directly proportional to the weight by which it is stretched. The natural frequency of a stretched string is given by the formula $n = 1/2l\sqrt{T/m}$ where n is the number of cycles per second, l the length of string in cm, T the tension in dynes and m the mass of string per cm of length.

A string when plucked or set into rapid vibration in any way very rarely vibrates with one motion through its entire length, it also vibrates in sections or *loops* which vary in number depending upon the length, tension and mass of the string. The fixed points separating the vibrating sections are called *nodes* (fig 77 10). If the string vibrates in halves it gives out not only its fundamental but its octave as well, if it vibrates in thirds the third harmonic is added, and so on. A column of air as in an organ pipe behaves in a somewhat similar manner, giving out a note composed of the

fundamental and a number of overtones. The different motions of the string or air column are imparted to the surrounding air with the production of waves, each of which is composed of the waves caused by the overtones superimposed upon the one due to the fundamental. The composite wave repeats itself at the frequency of the fundamental which therefore determines the pitch of the sound (see Complex Sounds below).

The *diatonic scale* comprises eight notes designated C D E F G A B C', including the fundamental and the octave. C is the fundamental or *unison* and C' the octave. Whatever the actual level at which the scale is pitched⁶ the ratios between the frequency of the fundamental and that of the other tones is constant. The seven notes D E F G A B C' can be produced by dividing a stretched string into 8/9, 4/5, 3/4, 2/3, 3/5 and 1/2 respectively. The frequency of each tone is the reciprocal of the corresponding fraction into which the string is divided, namely, 9/8, 5/4, 4/3, etc., etc. It follows that each tone in the scale also bears a definite ratio to the one below it. The frequency ratios are given below.

Note	C	D	E	F	G	A	B	C'
Ratio to C		9	5	4	3	5	15	
	1	8	4	3	2	3	8	2
Ratio to preceding note	9	10	16	9	10	9	16	
	8	9	15	8	9	8	15	

Two tones having a frequency ratio of $\frac{9}{8}$ or $\frac{10}{9}$ are said to differ in pitch by a whole tone. Those with ratios of $\frac{9}{8}$ are called *major tones* and those of $\frac{10}{9}$ *minor tones*.

If the ratio is $\frac{16}{15}$ the difference in pitch is a *half tone* or *semitone*.

The intervals are named according to their positions in relation to the first note of the scale, as follows:

Ratio	Name of interval
1 1	the unison
1 2	the octave
2 3	the fifth
3 4	the fourth
4 5	the major third

In the so-called *tempered* diatonic scale, to which tuning-forks, the pianoforte and certain other musical instruments are tuned, the octave is divided into twelve equal intervals. Each note rises in the scale by a semitone having a frequency exactly 1.05946 times greater than the one below it (1.05946 is the 12th root of 2, the ratio of the octave).

⁶ By international agreement the frequency of middle C of the piano has been set at 261 c.p.s.

QUALITY OR TIMBRE That property of sound by which one distinguishes between two tones of the same pitch and intensity, e.g., the note of a violin from a bugle note, is called *quality* or *timbre*. It is determined by the wave form. Sounds can be classified into *noises* and *musical tones*. Noises are defined as sounds (usually disagreeable or at least undesirable) possessing no regular period or definite pitch.⁷ Musical tones are due to waves which are repeated in regular sequence, they may be simple or complex, harmonious or discordant. A simple tone such as that produced by a tuning-fork is usually agreeable but monotonous and uninteresting.

The sound wave of a complex tone is a composite one formed by at least one wave of higher frequency superimposed upon that of the fundamental or simple tone. To the higher frequencies, i.e., to the overtones or harmonics, the particular quality of a given musical sound is due. The higher frequencies and the characteristic form of the wave for which they are responsible are produced within the wave length of the lowest or fundamental tone. Thus if the fundamental tone is 100 c.p.s. and waves of higher frequency—200 c.p.s., 300 c.p.s., etc.—are superimposed upon it, then the entire compound wave repeats itself 100 times per second, the pitch of the tone is determined by the latter frequency and its quality by the frequencies and volumes of the overtones (see fig. 77.11). The ear of the ordinary person cannot readily distinguish the separate overtones, but they are perceived by the trained ear of the musician. A complex sound can be analyzed into its components by means of a series of resonators, when the tone is produced in their vicinity only those resonators respond which have natural frequencies corresponding to the simple tone components. It was shown mathematically by Fourier that any regular periodic vibration can be resolved into two or more simple harmonic motions, the fundamental in all instances having the same frequency as that of the compound wave. This statement is referred to as Fourier's theorem.

The most harmonious sounds are complex tones, especially the combinations of such simple tones as the octaves C and C' (ratio 1 to 2), C and G (ratio 3 to 2) or C, E and G. Any tones whose frequencies are proportional to the simple numbers 1, 2, 3, 4, 5, 6 combine to produce an agreeable sound. Certain other combinations are discordant. According to Helmholtz's theory of harmony, the discordance is due to the production of beats (see p. 1196).

⁷ There are many discordant sounds possessing a regular period of vibration and a definite pitch which though technically classed as musical sounds would certainly be called noises by most persons.

AUDITORY SENSATIONS THE MECHANISM OF HEARING THE TELEPHONE AND RESONANCE THEORIES

THE THRESHOLD OF HEARING The least perceptible sound causes a pressure variation at the ear of $\frac{1}{12000}$ bar¹ or a force of $\frac{1}{12000}$ dyne per square centimeter, or in energy units, about 10^{-9} to 16^{-9} microwatts per square centimeter. These almost incredibly low values are for a sound vibration of around 2700 cycles per second but, as already stated, the level of the threshold of hearing is variable, depending upon the frequency of the vibrations. The ear is most sensitive to pitches ranging from 2000 c p s to 5000 c p s, i.e., to the upper two octaves of the pianoforte. The maximum sensitivity is for tones of 2700 c p s. Below and above the range from 2000 c p s to 5000 c p s the threshold rises rather rapidly (see fig 78.1). The lowest audible frequency is about 16 c p s, the highest between 20,000 and 30,000 c p s. At the upper and lower limits the intensity must be increased enormously above that required for a tone of 2700 c p s, a sound of the lowest frequency in order to be heard must have a pressure of 100 bars and at 20,000 c p s a pressure of 500 bars is required. In terms of energy, that required to make the highest frequency audible is nearly forty million million times greater than is necessary for a note of 2700 c p s.

The range of audible frequencies varies considerably between different species. The upper limit is highest in bats, sounds with frequencies far above the audible range of the human ear, namely 98,000 c p s can be heard. It is by means of the high-pitched (supersonic) cries which they emit and the detection of the echo of such sounds from objects in their path during flight that the bat is guided and enabled to avoid collisions.

THE THRESHOLD OF FEELING When the sound is very loud (at pressure variations above from 10 to 1000 bars, depending upon the frequency) it is

¹ The bar is a unit of pressure and amounts to about one millionth of an atmosphere or to a pressure of 1 dyne per sq cm. At sea level a decrease in pressure of 1 bar results from a rise in height of 8 mm. At the threshold of audibility for frequencies to which the ear is most sensitive the pressure variation is equivalent to a rise of only about one four hundred millionth part of a millimeter. At such a minute pressure the linear displacement of the membrane has a value comparable with molecular dimensions, or about $1/100,000$ that of the wave-length of green light.

felt as well as heard. The threshold of feeling is highest for frequencies between 250 c p s and 1000 c p s, and lowest for those near the upper limits of audibility. Thus a low rumbling sound is felt more than heard, and a high-pitched shrill note of high intensity arouses a decidedly unpleasant feeling within the ear. For most frequencies pressure variations above 600 bars cause pain and may result in damage to the auditory mechanism. The pain threshold is somewhat lower (between 100 and 200 bars) for frequencies between 2000 c p s and 5000 c p s, i.e., for frequencies with the lowest audible thresholds.

The discrimination of differences in intensity and pitch. The least perceptible difference in intensity (ΔI) is dependent upon the original intensity level and also upon the pitch of the sound. For sounds of ordinary intensity and frequency, e.g., tones of musical instruments and of the human voice, a difference of about 25 per cent is just perceptible. This represents a rise or fall of one decibel. It is only within this range of audibility that the Weber-Fechner law is even approximately obeyed. The percentage increase in intensity that is just perceptible is less than 25 at higher and greater than 25 per cent at lower intensity levels. For loud sounds (60 db or more) the least perceptible intensity difference remains fairly constant at between 5 and 10 per cent over a very wide range of frequencies, whereas a tone near the threshold of audibility and with a frequency of 60 cycles per second must be increased some 700 per cent before any difference is perceived. At low intensity levels $\Delta I/I$ varies with the frequency, it is smallest at 2050 c p s but increases progressively as the pitch is raised or lowered above this frequency.

The sensitivity of the ear for pitch discrimination is greatest over the range from 500 to 4000 c p s. The least perceptible difference in frequency (ΔF) varies very greatly in different parts of the scale. Within the range from 500 c p s to 4000 c p s $\Delta F/F$ has a value of 0.003, that is, a change in frequency from 1000 to 1003, from 2000 to 2006 or from 3000 to 3009 and so on, can be detected by the average ear. The trained ear of the musician gives a still lower value. At frequencies near the lower level of audibility $\Delta F/F$ is about 0.01 and near the upper level (16,000 c p s to 20,000 c p s) tones differing very widely in frequency can scarcely be distinguished from one another. Generally speaking,

pitch discrimination becomes less acute as the intensity of the sound is reduced.

When the sound intensity is kept at a constant medium value, and pitch discrimination determined over the whole audible scale, there are found to be some 1500 just perceptible differences or steps. When a sound of a constant medium pitch is gradually increased in intensity from the threshold of audibility the number of just perceptible differences in loudness is 325. One might expect that the product of these two figures would give the total number of tones which the ear can distinguish. But, owing to the fact that intensity discrimination varies with the pitch and vice versa, the number of distinguishable tones is only 320,000, not $(1500 \times 325 =) 487,500$.

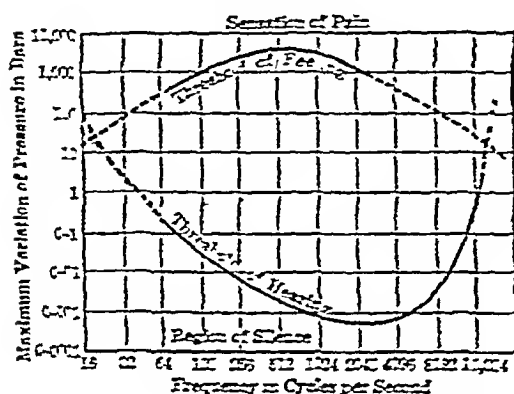


FIG. 73d. Chart showing the thresholds of hearing and feeling of sound at different frequencies (after Wegel, slightly modified).

MASKING. It is a common experience that a loud sound "drowns" a weaker one. We find it difficult to hear ordinary conversation in the roar of traffic, whereas a whisper seems loud in a quiet corner. It is generally true that for any given intensity low-pitched tones have a greater masking effect than those of high pitch; that is, a high-pitched tone is more readily masked by one of low frequency than vice versa. It is not, however, the absolute frequency of the masking tone which is of so much importance, but rather the relative frequencies of the two tones, the greatest masking effect resulting when they are of nearly the same frequency. For example, a greater masking effect is exerted upon a tone of 1000 cps. by a tone of 1050 cps. than by one of 830 cps. The phenomenon of masking offers a convenient method of measuring noise. A sound generated electrically by an audiometer is increased gradually until it is just perceptible to a person of normal hearing in noisy surroundings. Should it be necessary, for

example, to increase the intensity level of the audiometer tone to 50 decibels (i.e., 100,000 times) this value will be the measure of the masking or deafening effect of the noise. Masking is probably responsible for the phenomenon that persons suffering from middle ear deafness frequently hear conversation better in noisy surroundings. In this type of deafness low-pitched tones are usually affected to the greatest extent, and ordinary noise is largely made up of lower frequencies. The normal person, experiencing the effect of the noise from his own hearing, unconsciously raises his voice which is therefore heard more distinctly by the deaf person (see also p. 1211).

The masking of one sound by another is attributed, upon the basis of the resonance theory (p. 1205), to the overlapping of the vibrations of the two tones upon the basilar membrane. But a tone in one ear is also masked by a louder tone in the other, though to a less degree than when both tones are received by the same ear, this effect must, of course, be of central origin.

The principle of masking is utilized in certain hearing tests. In testing the hearing of one ear a sound which is capable of masking the test tone is applied to the opposite ear by means of an ear-telephone. The possibility of the test tone being heard by this ear rather than by the one under examination is in this way excluded. The masking sound will of course raise the threshold in the affected ear.

THE INTERFERENCE OF SOUND WAVES, BEATS

When two tones of nearly the same frequency are sounded simultaneously they meet in opposite phases (condensations and rarefactions of the air) at one instant and in the same phase at the next. As a result of the alternate interference and reinforcement, the sound undergoes a corresponding waning and waxing. At a certain periodicity it gives rise to a disagreeable throbbing sensation analogous to flicker in vision. Each period of maximum loudness is referred to as a *beat*. According to Helmholtz's theory of dissonance, the jangling or clashing quality of discord which results when certain tones are sounded simultaneously is due to beats produced by interference between the fundamentals or the overtones of the different tones in combination.

The number of beats occurring per second corresponds to the difference between the frequencies of the two tones. For example, if one tone has a frequency of 1000 cycles per second and the other 1100, the rate of beating will be 100 per second. If the frequencies

of the two tones are nearly the same and the beats therefore recur at relatively long intervals—2 or 3 per second—they are heard very distinctly but the sensation is as a rule not unpleasant, and at the slow rate of about 4 per minute they are usually not perceived Also, like flicker, the beats disappear if they occur very rapidly, i e, if there is a great disparity between the frequencies of the two tones, a harmonious blend then results or a new tone called a *difference tone* (p 1200) is heard At a certain rate of beating between these two extremes the discordance is maximal The frequency at which the beats disappear, as well as that at which the sensation is most unpleasant, depends upon the pitch of the lower tone, as shown in table 101 which gives the results of Mayer's experiments

TABLE 101

FREQUENCY OF LOWER TONE, CYCLES PER SECOND	NUMBER OF BEATS PER SECOND		INTERVAL BETWEEN TWO TONES AT WHICH BEATS DISAPPEAR
	At which beating most unpleasant	At which beats disappear	
			<i>semitones</i>
96	16	41	6
256	23	58	4
575	43	107	3
1707	84	210	2
2808	106	265	1 5

The sensation of beats in terms of the resonance theory (p 1205) is explained as being due to overlapping of the vibrations of two neighboring sections of the basilar membrane Thus the amplitude of vibration of a certain region of the membrane undergoes intermittent variations, corresponding alterations in the intensity of stimulation of the auditory receptors result. The ear, like the eye, is intolerant to this form of stimulation The greater the difference in frequency of the two tones the greater will be the distance between the two vibrating sections of the basilar membrane and the less tendency will there be for overlapping to occur and for beats to be produced One would expect then that the spread to resonators on either side of the ones in tune respectively with the two tones would be broader at high than at low intensities and that, as a result of the greater degree of overlapping, beats would be produced at smaller frequency differences in the former instance, such an effect of varying the intensity of the sound can be demonstrated On the other hand, two tones of nearly the same frequency which when sounded separately are inaudible, may as a result of overlapping give rise to audible beats when sounded simultaneously

The phenomenon of beats has an interesting application in the detection of fire damp in coal mines This gas, being lighter than air, transmits sound at greater velocity, a column of this gas has therefore a higher

vibration frequency than a column of air of the same dimensions (The vibration frequency of a column of gas is proportional to the time taken for the sound wave to pass through it) Two long tubes or whistles similar in every way are filled, one with the suspected mine air, the other with pure air The two whistles are blown simultaneously, if much fire damp is contained in the air of the first tube, the frequency of the sound vibrations will not be the same in it as in the one filled with pure air, and beats will be heard The concentration of the dangerous gas is estimated from the rate of beating

THE LOCALIZATION OF THE DIRECTION OF THE SOURCE OF A SOUND A visual sensation is projected very accurately to a definite point in the outside world, i e, an object in the visual field forms an image upon a corresponding part of the retina Sound "images" cannot, of course, be localized in the same way upon the auditory re-

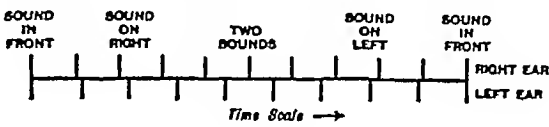


FIG 78.2

ceptors, there being no spatial correspondence between the sounding body and the basilar membrane In the judgment of the direction of a *single brief sound* or of an *intermittent sound* the most important factor is the difference in the arrival times of the sound at the two ears If the sound reaches both ears simultaneously, it is projected to the mid-line in front or behind the head, but if the vibrations are not received at the same instant by the two ears the sound is located on the side of that ear which is stimulated first An illusion of a sound changing its direction may be created by means of two watches which tick at slightly different rates held one to each ear Just as with any two movements of different frequencies, the time intervals between the ticks of the two watches lengthen and shorten periodically (fig 78.2) When the two ticks fall simultaneously upon both ears, a single sound is heard in the median plane either in front or behind An instant later when the ticks become asynchronous the sound moves to the side of the faster tick, but as the intervals between the sounds lengthen they are heard separately by each ear, then as the intervals shorten again a single sound is heard which now seems to come from the side of the watch with the slower tick Thus the

sound image seems to move continuously around the head

The effect upon sound localization of varying the time interval between the stimulation of the two ears can also be demonstrated in the following manner. A stethoscope in which the length of tubing of one limb can be altered (fig 78.3) receives sound vibrations from in front and in the mid-line. When the pathway is lengthened by extending the tubing of the adjustable limb the sound reaches the opposite ear first, and therefore seems to come from that side, upon shortening the tubing the sound is located on the side of the shorter limb.

The maximum difference in the arrival times of a sound in opposite ears at which the separate sensations fuse into one is determined by the length of the path from ear to ear, namely, 21 cm (directly through the head). Sound is transmitted this distance through air in 0.00063 second and through the longer distance around the obstruction caused by the head in a somewhat longer time. The greatest interval between the arrival of a single sound at each ear is therefore a little longer than 0.00063 second, and this is the maximum interval between two sounds which when led to separate ears should be heard as one. By actual measurement it has been found that at intervals greater than about 0.00180 the two sounds are heard separately. The ear through experience discounts a time interval caused by the difference in distance of the two ears from the origin of the sound but recognizes any greater time interval as being due to separate sounds.

The localization by the ear of a *continuous sound* cannot be explained by an interval between the reception of the sound by the two ears. When the vibration frequency is less than 800 c.p.s. a difference in phase of the sound waves striking opposite ears is the most important factor. Waves emitted from a source in the mid-line reach the two ears simultaneously and the sound is localized accordingly, but if the sound comes from one side the crest of the wave will reach the ear of that side an instant before it reaches the opposite ear. When the length of the sound wave is double the distance between the ears, namely, 42 cm, the waves fall upon the two ears in opposite phases, one wave cannot then be said to be in advance or behind the other. The ear cannot distinguish between the two phases, the sound is localized with difficulty, seeming to come from all sides. The uncertainty in localization becomes evident at around 800 c.p.s. (length of wave about 42 cm). At frequencies above from 800 c.p.s. to 1000 c.p.s.

localization by phase difference fails. High-pitched sounds such as the chirp of a cricket are very difficult to locate and often seem to come from several directions at once. Some information as to the direction of high tones is gained, however, from *differences in the intensity* or in the *quality* of the sound in the two ears. Simple tones of low pitch (frequency less than about 1000 c.p.s.) cannot be located through a difference in intensity in opposite ears owing to the great length of the waves (measured in feet or meters) as compared with the dimensions of the head. That is to say, the head casts no shadow for low-pitched sounds. On the contrary, sounds of high frequency (short wave length) coming from one side are shielded

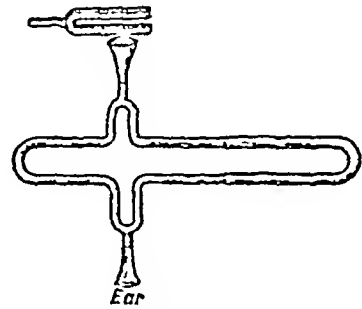


FIG 78.3

from the opposite ear by the head, and the difference in intensity of the vibrations in the two ears is used as a clue to the direction of the sound. The shielding effect of the skull also plays a rôle in the localization of a complex sound situated on one side or behind, even though it is of low pitch, because the higher frequencies (overtones) are screened by the skull and the auricle, the alteration in the quality of the sound caused thereby then serves as a clue. This factor is important particularly in deciding the direction of familiar sounds, i.e., those, such as voices, which have been heard upon previous occasions from different angles and hence have had their qualities impressed upon the memory.

In the localization of sounds with frequencies below 1000 c.p.s. it would obviously be an advantage to have the two receptor organs separated by a greater distance than 21 cm, for any phase difference would be increased thereby. In World War I this was recognized, an apparatus consisting of two microphones separated by a difference of about two feet and connected by tubing to the ears of an observer was employed for locating enemy aircraft in flight. The

application of the principle is also seen in the insect world, certain forms have paired sets of auditory receptors separated as widely as possible from one another, namely, on either side of the thorax or abdomen or, as in crickets and grasshoppers, just below the knee joints

THE MECHANISM OF HEARING

THE EXTERNAL EAR

Though the shape of the human auricle suggests that it might serve as a concave surface to concentrate sound vibrations and direct them into the auditory meatus, its small area as compared with the length of the waves of ordinary sound, negatives the idea that it acts in this manner to any important degree. Only sounds of relatively high frequency (over 6000 c.p.s.) whose wave lengths are commensurate with the size of the auricle could be funnelled in this way, longer waves would tend to be scattered. The large trumpet-shaped ears of some animals, on the other hand, are of such a size as to be of real service in this respect, especially since they can be turned towards the source of the sound.

THE TYMPANIC MEMBRANE is aperiodic. That is, it vibrates unselectively to a wide range of frequencies, having no natural frequency of vibration, or rather, its natural frequency lies below the threshold of hearing. Like the well-designed diaphragm of a telephone transmitter, it does not resonate to any particular frequency but responds to all. It is forced into vibration by the sound waves and exhibits a high degree of damping, its movements stopping almost instantly upon cessation of the sound.

The tympanic membrane does not faithfully reproduce the *form* of the incident sound waves, but modifies them, the new vibratory motion being then transmitted through the ossicles to the internal ear. This behavior of the membrane is due to its asymmetry. A structure, such as the prong of a tuning-fork or the skin of an ordinary drum which is perfectly elastic and symmetrically loaded moves equal distances on either side of its resting position, its movements in response to a pure tone being simple harmonic and can be represented by a sine curve. The asymmetry of the ear drum is due mainly to the auditory ossicles to which it is connected on its inner aspect, and to the tension exerted by the intra-aural muscles.

TONES CREATED BY THE EAR, AURAL HARMONICS, COMBINATION TONES. The asymmetry of the drum membrane results in the production

of tones by the ear itself, i.e., additional tones with frequencies differing from those of the incident sound. The curve in fig 78.4 represents the movements of the drum membrane to a pure tone of 100 c.p.s. Upon analysis this curve is found to be compounded of a series of simple harmonic waves having frequencies of 200 c.p.s., 400 c.p.s., 600 c.p.s., etc. The compound wave repeats itself 100 times per second, so the original tone as well as the higher frequencies is heard. Thus, as a result of the asymmetry of the drum membrane the ear has created a new set of frequencies, adding to the original tone the octave (200 c.p.s.) and the other harmonics. These overtones, called *aural harmonics*, being purely subjective, i.e., non-existent outside the ear, cannot, of course, be detected by means of resonators.



FIG 78.4 The response of the ear-drum to a simple sound. The unsymmetrical nature of the human ear results in parts of the simple harmonic curve, which is drawn thick, being replaced by those shown by broken lines (After Jeans)

The aural harmonics are not heard, especially those of higher frequency, unless the original sounds are fairly loud (over 45 decibels) for it is only when the vibrations reach a certain amplitude, and force the drum membrane and the structures of the middle ear to their elastic limits that they become asymmetrical. As the intensity of the sound is progressively increased harmonics of higher and higher frequencies are added to the original tone. The latter, however, always tends to obscure (mask) to a certain degree the higher frequencies. With vibrations of small amplitude, movement at the incudomalleolar joint permits equal displacements of the membrane in both directions. Also, when this joint is ankylosed, as in old age, outward and inward movements tend to become more nearly equal and the subjective tones are absent or greatly reduced. The tensor tympani muscle, by drawing the malleus inwards, tends to reduce the movement at the incudomalleolar joint. Increased tension of this muscle therefore tends to increase the aural harmonics, relaxation to diminish them.

Combination tones. When two pure tones are sounded, the ear adds not only the octave of each tone, but a *difference tone*, so called because its frequency is the exact difference between the frequencies of the

two primary tones. For example, when tones with frequencies of 200 and 300 cycles per second are sounded together a compound motion will be produced which repeats itself 100 times per second. Tones of 100, 200, 300, 400, 500, 600 c.p.s., etc., will be heard. The frequencies 200 c.p.s. and 300 c.p.s. are, of course, those of the two primary tones and 400 c.p.s. and 600 c.p.s. are their octaves, 100 c.p.s. is the "difference" tone and 500 c.p.s. (the sum of the original frequencies) is called the "summation" tone.² The summation tone is always more difficult to hear than the difference tone since, being of higher frequency and weaker, it tends to be masked by the primary and difference tones.

The difference tone of 100 c.p.s. in the foregoing example is the fundamental of the tones 200 c.p.s. and 300 c.p.s., and it is generally true that whenever two or more tones which happen to be the harmonics of the same fundamental tone are sounded together

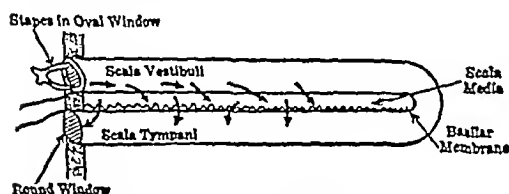


FIG 78.5 Diagram of the passages of the cochlea straightened out to show the manner in which vibrations are transmitted from the oval to the round window through the scala media. The wavy line represents the organ of Corti.

with sufficient intensity, the fundamental is heard as a difference tone, i.e., the fundamental tone is created by the ear. To give another example, when the second (octave) and the third harmonics of middle C are sounded simultaneously (respective frequencies are 512 c.p.s. and 768 c.p.s.) middle C itself (256 c.p.s.) is heard.

These facts have important practical applications in telephone, talking picture and radio engineering. The diaphragm of a telephone is so constructed that it responds to the harmonics of the human voice but not to the main tones, these are added by the ear of the listener. The natural frequencies of the instrument range from 300 to 2400 c.p.s., which are above the frequencies of the main vocal tones. The latter are therefore transmitted at a scarcely audible intensity, whereas the higher frequencies—the harmonics—since they coincide with the natural frequencies of the instrument are strengthened. The deeper main tones

² Difference and summation tones are not always subjective, for under certain conditions they are produced outside the ear, as when the sounds are transmitted through some medium which like the tympanic membrane is asymmetrical, e.g., the rectifying valve of a radio set.

are created by the ear. By this device exaggeration of certain of the more prominent tones of the voice is avoided. Otherwise these tones through resonance would mask other tones of higher frequencies and lower intensity which are largely responsible for the quality and intelligibility of speech. The principle has been demonstrated in a striking manner in the laboratories of the Bell Telephone Company. Two sets of gramophone records were made. A song, a conversation, instrumental music and other sounds were reproduced on one set, all the tones being faithfully recorded. The other set of records were made from the same sounds after they had been passed through filters and the fundamental tones removed. To the ear both sets of records were almost identical.

THE TRANSMISSION OF THE SOUND FROM THE DRUM MEMBRANE TO THE INTERNAL EAR

The vibrations of the tympanic membrane set up by air-borne sounds are transmitted to the oval window by the chain of auditory ossicles and thence to the perilymph of the vestibule. Vibrations of audible frequencies are transmitted through Reissner's membrane and the scala media to the basilar membrane, then through the scala tympani to the round window (fig 78.5). The compression phases of the sound waves in the external auditory meatus cause a movement inwards of the tympanic membrane and of the base of the stapes, a downward displacement of the basilar membrane and a movement outwards (i.e., towards the tympanum) of the round window. During the expansion phases of the sound waves, displacements of these structures in the reverse direction occur. Thus are the vibrations of the drum membrane reproduced in the basilar membrane.

It is quite evident that, since liquids are incompressible, no appreciable movement of the stapes could occur were there no part of the bony labyrinth which yielded to pressure. The membrane of the round window situated between the scala tympani and the middle ear yields readily, and thus permits such a movement to take place. Owing to the presence of the helicotrema through which the scala vestibuli communicates with the scala tympani, vibrations of very low frequency are incapable of setting up vibrations in the membranous cochlea of sufficient force to stimulate the nerve endings, the fluid is moved *en masse* as in a U-shaped tube formed by the upper and lower galleries.

The middle ear is filled with air, vibrations of the drum membrane must therefore be transmitted

to the round window as well as to the oval window. The auditory ossicles, however, provide a preferential path for the transmission of sound, serving as a mechanism for concentrating the vibrations at the internal ear. Any inward movements of the membrane of the round window which may result, since they will be opposite in phase to those of the base of the stapes, will tend to diminish rather than to enhance the force of the vibrations of the basilar membrane. It has been found, for example, that shielding the round window by means of a pledget of cotton placed over the niche in which it lies increases the acuity of hearing.

Two factors are responsible for the intensification of the vibrations at the oval window, (a) the leverage action of the malleus and incus and, (b) the greater area of the drum membrane as compared with that of the oval window. The malleus and incus together form a bent lever of the first class, the handle of the malleus constituting the long arm of the lever and the long process of the incus, which articulates with the stapes, the short arm (see fig 78.6). The axis of rotation (fulcrum) of the lever is represented by a line running through the anterior ligament of the malleus and the short process of the incus. The length of the short arm is about $\frac{1}{3}$ that of the long arm formed by the handle of the malleus. The amplitude of movement of the tip of the incudal arm is therefore only $\frac{1}{3}$ that of the tip of the long arm, i.e., the amplitude of movement at the center of the drum membrane. The matter is complicated by the fact that the base of the stapes is not moved in and out in a simple push-pull or plunger-like action but, being hinged at its posterior margin³ executes a rocking motion, the amplitude of the movement of the posterior part of the footplate is considerably less, therefore, and that of its anterior edge considerably greater than $\frac{1}{3}$ that of the drum membrane. Should the sound be excessively loud (i.e., near the threshold for pain) the footplate, instead of moving as just described, pivots around a longitudinal axis passing through its center (Békésy). The delicate structures of the internal ear are further protected from violent vibrations by the momentary dislocation of the incudomalleolar articulation. Instead of engaging in the usual way the surfaces glide over one another, like a slipping motor clutch, with consequent loss of energy and reduc-

tion in the amplitude and force of the vibrations. The protective rôle played by the intra-aural muscles has been mentioned (p. 1186).

The ratio of the long arm of the lever to the short arm being 3 to 2, one would expect an increase of $\frac{3}{2}$ in the force delivered at the oval window over that incident upon the tympanic membrane. Actually, the energy lost in transmission (in overcoming the inertia of the ossicles) and the damping effect of the surrounding air more than offset the mechanical advantage gained by leverage. Were no other factor involved the vibrations would have a force only half as great at the oval window as at the drum membrane. But the area of the latter is some 20 times greater than that of the oval window. This alone by

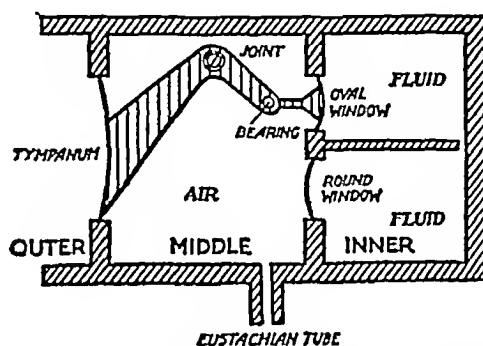


FIG 78.6 Model illustrating the transmission of sound vibrations from the drum membrane to the organ of Corti. (From Lythgoe after Beatty.)

concentrating the sound waves would increase their force upon the smaller area twenty-fold. But the loss in transmission, as just stated, is about 50 per cent, which means that the force incident upon the drum membrane is about $\frac{1}{10}$ of that upon the oval window. This increase in force of the vibrations with relatively little change in amplitude is of considerable importance. It appears that a change of just such a nature is required for the most efficient transference of vibrations from a medium of low resistance, namely air, to one of greater density and high resistance, such as the fluid of the internal ear.

BONE CONDUCTION The auditory receptors are stimulated by sound transmitted through the bones of the skull. In man the most direct and effective bony pathways for the conduction of sound to the internal ear is furnished by the trabeculae in the sub-occipital region. Injury to this region, as by fracture, results in impairment of bone conduction. In order for sound to be perceived to any important extent through bone the sounding body must be in contact

³ According to Wilkinson, the base of the stapes pivots around an axis passing through the junction of its posterior and middle thirds.

with the skull (e.g., mastoid process), bone conduction of air borne sounds impinging upon the head is very slight. The sound is transmitted through the bone as through any solid, as a train of alternating compressions and expansions. Owing to the presence of the oval and round windows, chiefly the latter, serving as outlets for pressure variations, adequate movements of the basilar membrane are permitted to take place. When the base of the footplate is immovably fixed in the oval window, the round window alone serves such a purpose. The waves of compression, according to Békésy, exert their effect mainly upon the semicircular canals from which fluid is forced into the vestibule. For this reason, and also owing to the greater resistance offered at the oval window, the pressure in the scala vestibuli is elevated more than that in the scala tympani. With each wave of compression the basilar membrane therefore moves downwards and the membrane of the round window outwards.

In a person with normal hearing, obstructing the external auditory canal by the finger increases the loudness of bone conducted sound. This is attributed to the compression of the air in the canal by the vibrations through the skull, the sound being thus transmitted through the ossicles to the internal ear as in ordinary air transmission. The threshold for bone conduction is given by Békésy as 3.5×10^{-9} cm vibratory movement of the skull for a tone of 800 c p s.

STIMULATION OF THE HAIR CELLS The tips of the hair cells are embedded in the under surface of the tectorial membrane. An upward movement of the basilar membrane should cause bending of these processes and compression of the bodies of the cells, a downward movement of the membrane would, on the other hand, cause traction upon the processes, and possibly some elongation of the bodies of the cells. It is not definitely known which of these movements is responsible for the stimulation of the nerve endings. The nature of the immediate stimulus is also a matter for conjecture, it may be mechanical, electrical (see cochlear potentials, p 1203) or chemical. A theory of chemical stimulation has been proposed by Hallowell Davis. He suggests that the compression of the sensory cells by the upward swing of the basilar membrane, i.e., when the base of the stapes moves outwards, causes the liberation of a chemical substance which, acting upon the sensory endings, sets up the nervous discharge.

THEORIES OF PITCH PERCEPTION

There has been relatively little room for dispute concerning the functions of the middle ear. The mechanism of the internal ear, on the contrary,

has been for many years a subject of lively controversy to which both physiologists and physicists have contributed. The appreciation of pitch has been the pivot around which discussions of cochlear functions have turned, and the several theories which have been proposed are attempts to explain this extraordinary faculty of the auditory mechanism.

All theories of hearing can be divided into two main categories. In one group (a) are placed those theories which attribute the estimation of pitch, and in consequence the ability to analyze a complex sound into its components, to the cerebral centers, and in the other (b) those which maintain that sound analysis is purely a function of the cochlea. None of those coming under the first heading is acceptable as a complete theory of hearing, one, namely the telephone theory, will be briefly described.

The telephone or frequency theory was formulated by Rutherford in 1886. The diaphragm of a telephone transmitter converts sound vibrations into electrical impulses (actually, rhythmical variations in current strength) of the same frequency. The electrical impulses conveyed by wire to the receiver of the instrument set up vibrations in a diaphragm which reproduces the original sound. If we substitute the basilar membrane for the diaphragm of the telephone, the electrical impulses for nerve impulses, the auditory nerve for the wire, and the auditory sensation for the reproduced sound in the receiver, we then have a rough outline of the telephone theory of hearing. In this theory the basilar membrane is supposed to vibrate as a whole, i.e., like the diaphragm of the telephone transmitter, not selectively according to the principle of resonance. The telephone theory postulates, therefore, that *the cochlea possesses no faculty of sound analysis, the frequency of the impulses transmitted over each fiber of the auditory nerve being the sole basis for the estimation of pitch.* The intensity of sound is, according to this theory, a function of the number of active nerve fibers.

The relatively low frequency (less than 1000 per second as compared with from 20,000 to 30,000 c p s for the highest audible pitch) at which the nerve fiber can conduct impulses has always been an insuperable objection to the telephone theory. This point will be discussed in the next section.

THE WEVER AND BRAY PHENOMENON

In 1930 Wever and Bray made the surprising discovery that spoken words and other sounds

entering the ear of a decerebrate animal were faithfully reproduced when electrodes were placed upon the auditory nerve and connected through a valve amplifier with a telephone receiver or loud speaker. It was at first thought that the sounds emitted by the telephone receiver were due solely to action currents (i.e., to nerve impulses) of the same frequency as the sound vibrations received by the animal's ear. If this were true it would mean, of course, that the quality of sensation is correlated with the frequency of impulses in the nervous discharge and would therefore be contrary to observations upon other senses, for all recent work indicates that impulse frequency varies only with the *intensity* of the stimulus. Such a conclusion implies also that the analysis of sound is a function of the cerebral cortex, as postulated by the telephone theory. As a result fresh interest in this theory was aroused.

Saul and Davis in later experiments demonstrated that there are two components in the electrical potentials picked up from the nerve, namely, *true action potentials* (nerve impulses) resulting from the excitation of auditory nerve endings in the cochlea, and electrical potentials generated within the cochlea as a result of the distortion of non-nervous structures by the sound waves. These latter effects are referred to as the *cochlear response*, *aural microphonics* or *cochlear potentials*. In records showing both types of electrical effect the action potential waves are superimposed upon the aural microphonics.

The cochlear response differs from the action potentials in the following respects. General anesthesia abolishes the action potentials but not the microphonic effect. The action potentials are also affected much more by cooling the cochlea, and they disappear much sooner after the death of the animal. Arrest of the blood supply to the cochlea rapidly abolishes both responses. The latent period of the action potential (up to 0.83 milliseconds) is much longer than that of the aural microphonics (0.1 milliseconds). The cochlear potentials can be recorded from any part of the internal ear or even from any part of the skull, provided that the amplification is adequate. (They are obtained most readily with the differentiated electrode at the round window and the indifferent electrode inserted into the muscles of the neck.) The action potentials can be recorded only from the auditory nerve or some part of the auditory pathway, e.g., cochlear nuclei of the medulla, lateral lemniscus, etc. The cochlear potentials have a much greater tendency to spread through the tissues, especially over meningeal surfaces, it is owing to such spread that they can be picked up from the

auditory nerve. The waves of the two responses differ in shape, the wave form of the cochlear potential being almost identical with that of the stimulating sound. The action potentials synchronize with the sound waves up to a maximum of 3000 per second (see p. 1204), whereas the cochlear potentials follow the vibration frequency up to 16,000 cycles per second, and it is probable that with more delicate methods of recording a correspondence between the cochlear potentials and the sound waves up to the limit of audible frequencies, namely, 20,000 or more cycles per second, could be demonstrated. Finally, the cochlear potentials show summation, whereas the action potentials are of course "all or none".

More than one theory as to the mode of origin of the cochlear potentials has been offered. Davis and his associates believe that they are due to pressure variations upon the hair cells induced by the sound waves, they therefore come under the head of *piezo-electric currents* such as are generated by pressure upon a quartz crystal. According to Davis, a potential difference is created between the bases and free ends of the hair cells when these cells are compressed against the overlying tectorial membrane, that is, when the footplate of the stapes moves *outwards* and the basilar membrane swings upwards. In support of his contention Davis cites experiments upon animals (albino cats and waltzing guinea-pigs) in which the organ of Corti was congenitally absent or the hair cells abnormal and no aural microphonics could be obtained. Moreover, certain chemicals, e.g., crystals of sodium chloride, applied to the internal ear injure the hair cells and depress the cochlear response, and when degenerative changes were induced in the mid-region of the basilar membrane of dogs by a loud tone of medium pitch (p. 1207) sounded over a long period, this particular sound was not heard by the animal nor was it followed by a cochlear response.

The physiological function, if any, of the cochlear potentials is unknown. They appear to be rather of an accidental occurrence—an epiphenomenon of auditory function. However, this much can be said, hearing may be lost though the aural microphonics persist unimpaired, but hearing is never retained in their absence.

Evidence that vibrations of resonant structures in the cochlea are responsible for the aural microphonics has been secured by Walzl and Bordley. They have succeeded in injuring selected and limited regions of the organ of Corti, without separation of Reissner's membrane on the basilar membrane, and have studied the effects of such injuries upon the cochlear potentials. Damage near the base of the cochlea was found to raise the threshold for high tones while injury near the

apex was followed by impairment or loss of the response to sounds of low frequency. These results point to elements—most probably the hair cells of the organ of Corti—attached to a particular level of the basilar membrane, and thus tuned to particular sound frequencies; they give strong support for a place theory of hearing (p 1207)

Impulses recorded from the auditory nerve and brain Action potentials free from microphonic effects can be picked up from the auditory nerve by means of coaxial electrodes (described on p 922). The latency of the action potentials is between 0.53 msec. for sounds of the highest intensity and 0.83 msec. near the threshold of audibility. The impulses are synchronous with the sound waves up to about 3000 c.p.s., above this frequency there is no correspondence between the two. When a sound with a frequency less than 900 c.p.s. is gradually increased in intensity a point is reached at which the action potential is maximal. If the intensity is kept at this level but the frequency raised above about 900 c.p.s., the amplitude of the waves falls sharply to about half of their previous value, at 2000 c.p.s. the amplitude is reduced to one-third and at 3000 c.p.s. or so, to a very low value. It must be remembered that the electrodes placed upon the intact nerve record potential changes from a number of fibers almost simultaneously. The recorded waves are, therefore, composite in nature and vary in amplitude with the number of functioning fibers. The step-like fall in amplitude as the frequency is raised from 900 to 3000 c.p.s. is due to a corresponding reduction in the number of active fibers and this reduction is the result, in turn, of the impulses falling each in the relative refractory period of its predecessor. Now the absolute refractory period of a fiber of the auditory nerve is about $\frac{1}{1000}$ second. In other words, a single fiber cannot conduct more than 1000 impulses per second. Therefore, when the sound vibrations reach a frequency of between 900 and 1000 c.p.s., the fibers respond only to every other stimulus. But the refractory periods of all the fibers do not begin and end at the same instant, this allows half the fibers to respond to alternate vibrations and the nerve as a whole to respond to every vibration though, as just stated, the amplitude of the action potential is reduced by half. At a vibration frequency of 2000 or 3000 c.p.s. the individual fibers respond to every third or fourth vibration, respectively. Yet again, since the refractory periods are

not coterminous, the fibers function in *rotation*, and the nerve as a whole is capable of answering to every stimulus, the amplitude of the action potential is diminished in proportion to the number of inactive fibers.

Thus, the limited correspondence of the auditory impulses with the frequency of the stimulating sound is a phenomenon which is due, apparently, to the fact that the auditory nerve is a composite record of the potentials in a number of separate nerve fibers. The fiber responds to a definite phase of the stimulating wave-cycle, but each fiber in a group does not necessarily respond to the same wave, so, the frequency of the impulses in a *single fiber* may be very much lower than that of the stimulating sound. Galambos and Davis have recorded the impulses from a single fiber by means of a glass microelectrode inserted into the auditory nerve. The nerve was exposed through an opening in the temporal bone and the electrode adjusted by a micromanipulator until a single series of impulses appeared in the record. The findings of Galambos and Davis are entirely in accord with a place theory of hearing (p 1207). The isolated auditory fiber was found to behave in the same manner as does any other sensory nerve. At threshold intensity it responded to a narrow band of sound frequencies, the impulse frequency varied only with changes in the *intensity* of the sound. The maximum frequency of the impulses, 400 per second, was recorded when the sound was increased to 30 decibels above threshold intensity. At the higher sound intensities the fiber reacted less *specifically*, responding to frequencies within a wider range. In some instances a very loud sound, even an octave below the band of frequencies to which the response of the fiber was restricted at threshold intensity, caused a discharge of impulses.

Impulses can be recorded by means of electrodes inserted at various levels along the auditory pathway, e.g., the cochlear nuclei, the trapezoid body, the lateral lemniscus, the medial geniculate body or the inferior colliculus, and even from the auditory radiation and auditory cortex. At the higher levels the frequency at which synchronization of impulses and sound waves ceases is much lower than in the auditory nerve. The number of synapses which the impulses must cross is apparently the factor responsible for depressing the synchronization limit. At the inferior colliculus synchronization is not evident, as a rule, above

frequencies of 1000 c p s and not above 20 c p s in the auditory radiation or cerebral cortex

Now the highest audible frequency is from 20,000 to 30,000 cycles per second. It is quite evident therefore that the telephone hypothesis or any other based upon the assumption that the brain receives impulses of the same frequency as that of the incident sound can no longer be entertained.

The organ of Corti is projected to the medial geniculate body, which constitutes the primary auditory center, and thence to the acoustic area of the cerebral cortex. In both these regions a spatial differentiation according to sound frequencies has been demonstrated. The dorsal part of the medial geniculate body responds to frequencies of around 8000 c p s, the anterior and lateral parts to 4000 and 2000 c p s, respectively, and the posterior part to 1000 c p s. Cycles of 500 per second and less are recorded in the central portion.

THE RESONANCE OR HARP THEORY

Though it was suggested as long ago as 1761 by Cotugno of Naples and in 1826 by Sir Charles Bell of Edinburgh that the ear owed its faculty of pitch perception to resonating structures in the labyrinth, it was not until 1863 that the scientific foundations of the resonance theory were laid. In this year Helmholtz published his now famous work entitled *Sensations of tone as a physiological basis for the theory of music*, in which he gives an account of a brilliant series of physical and physiological studies of hearing.

The resonance theory postulates that the *analysis of sound into its constituent frequencies is primarily a function of the cochlea*, that the fibers of the basilar membrane constitute a series of resonators, and that the part or level of the membrane at which the fibers are set into maximal vibration by the sound waves is the sole basis upon which the brain rests its judgment in discriminating differences of pitch.

The fibers of the basilar membrane which in deference to the resonance theory are frequently referred to as the *auditory strings*, number about 24,000. Each arch of Corti, of which there are some 5000, is associated with 4 or 5 auditory strings. The combined structure, rather than each one of the 24,000 fibers, should probably be regarded as a resonating unit. In this theory the resonating structures are compared to the strings

of a harp or pianoforte, they are believed to vibrate in sympathy with the sound waves, i e, each resonator responds to the frequency corresponding to that of its own free vibration (p 1191). We have seen (p 1188) that the fibers increase progressively in length from the base to the apex of the cochlea, a fact quite in accord with the resonance theory.

Upon first thoughts one would likely conclude that this resemblance of the basilar membrane to a stringed instrument must be mere coincidence and that it is rather fantastic to compare a structure only 30 mm long and 0.5 mm wide at its broadest part with a piano or harp in which the longest (bass) wires are measured in feet. Furthermore, the auditory range is some 11 octaves and some 1500 separate tones can be distinguished by the ear, whereas the pianoforte extends over only 8 octaves, containing some 96 separate notes. But factors other than length, namely, tension and mass (p 1193) determine the vibration frequency of a stretched string. There is reason to believe that the basilar fibers are differentiated as to tension, the loosely stretched fibers being at the apex where the flimsy character of the external spiral ligament seems to indicate that it exerts here the minimum degree of traction upon the basilar membrane. In the basal turn the external spiral ligament is several times thicker than at the apex, which suggests that in this part of the cochlea the basilar fibers are at their greatest tension. Yet, even granting that the longest auditory strings are at minimum tension, their mass would need to be greatly increased in order that they should be capable of vibrating at the lowest audible frequency, namely, 16 cycles per second. Now, the bass strings of the piano are, of course, of greater length and lower tension than the strings of the treble, but their low frequency of vibration is secured largely by increasing their mass. The greater mass is obtained not so much by increasing the thickness of the wire (for this would greatly reduce its flexibility) but by a helix of copper wire, i e, by a wire coiled around it. By means of such loading of a comparatively short thin wire, a vibrating element is obtained with a low frequency yet sufficiently flexible and of convenient length.

The auditory strings are loaded by the cellular structures of the basilar membrane (see p 1188) and also, to a greater extent, by the fluid in which they are immersed. This liquid load, according

to Wilkinson, is the column of fluid extending upwards from the oval and round windows to the basilar fibers resonating at the moment. The length of the fluid column will vary, of course, with the level of the vibrating fibers, i.e., with their distance from the oval and round windows. According to this conception of liquid loading the auditory strings are differentiated with respect to mass as well as according to length and tension, the shortest fibers at the base of the cochlea (i.e., nearer the oval and round windows) being loaded with the shortest liquid columns, the fibers at the apex with the longest columns. This principle of liquid loading is illustrated in fig. 787.

In order to account for the very high frequency (20,000 to 30,000 c.p.s.) of the auditory strings at the base of the cochlea, one must suppose that

they could not act as a series of resonators. This apparently insuperable objection to the resonance theory has been met by the conception of *maximum stimulation* advanced by Gray. For example, a sound sets into vibration not only those basilar fibers which are in tune with it, but also, though less forcibly, those for a variable distance on either side. This assumption is in agreement with observations upon pitch perception after auditory fatigue (p. 1192). It follows that there will be a central fiber or small group of fibers which vibrates maximally and other "out of tune" fibers which vibrate submaximally, the amplitude of vibration decreasing progressively with the distance from the "in tune" or central resonators. Only those vibrations of maximal amplitude, it is claimed, lead to effective stimulation of the nerve endings.

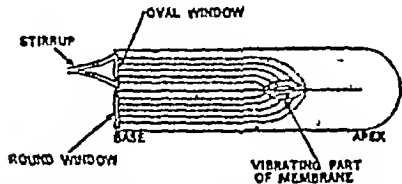


FIG. 787. Illustrating liquid loading of the auditory strings. See text. (After Beatty.)

the fibers in this situation are under a tension which, to opponents of the resonance theory, has seemed incredibly high. The length of the shortest basilar fiber is in the neighborhood of 0.16 mm. Such a fiber in order to be in tune with the highest audible frequency would need to be under a tension of nearly 4 tons per square inch (Beatty). Yet such a stretching force, enormous as it is, does not exceed the bounds of possibility when the strength of other biological materials is considered. A human hair, for example, can sustain a stress of 9 tons per square inch, a spider's thread, 12 tons per square inch and silkworm gut, 32 tons per square inch.

The basilar fibers are embedded in a homogeneous ground substance which by binding them into a continuous structure prevents them from vibrating separately, for this reason it would seem that

"The idea of a liquid load upon the basilar fibers was stated by Helmholtz in the following words: 'That such short strings should be capable of corresponding with such deep tones must be explained by their being loaded in the basilar membrane with all kinds of solid formations, the fluid in both galleries must also be considered as weighting the membrane, because it cannot move without a kind of wave motion in the fluid.'"

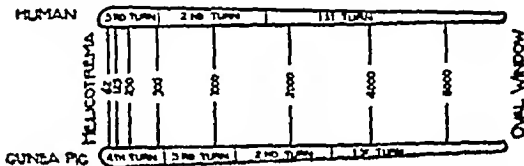


FIG. 788. Shows relative positions of the tones on the basilar membrane in the human subject and in the guinea pig. (After Stephens, Davis and Lune.)

This concept of maximal stimulation is not merely an ingenious hypothesis, for the same principle appears to be the basis of cutaneous localization. When a pointed object is pressed upon the skin, e.g., on the finger tip, the sensation is restricted to a single small area, even with deep pressure one feels only a sharp point. Yet the skin for an area of several square millimeters surrounding the point of maximal stimulation is deformed sufficiently to arouse a definite sensation were the stronger stimulus absent.

It is also pointed out by Wilkinson that in order for the liquid columns to load the membrane in the manner described, it is necessary that the fibers be joined along their lengths by a material impervious to the fluid, otherwise eddies set up between the individual strings would cause vibrations differing in frequency from those of the incident sound waves.

Though the account given in the foregoing paragraphs offers an interesting and reasonable explanation of cochlear function, direct evidence is not sufficient to justify one in expressing dogmatic views upon the complex problem. Nevertheless certain points seem to be definitely established.

in the *first* place, pitch perception is a function of the cochlea and does not depend, as claimed by the telephone theory, upon the frequency of the impulses received by the auditory center. *Secondly*, this membrane contains structures which relate vibrations to some definite cochlear level according to their frequencies, and the excitation of nerve endings at a particular level is the basis upon which the brain discriminates between differences of pitch. A diagram representing the levels of the cochlea responsive to various frequencies is shown in fig 78.8 (It will be noted that the lower octaves are crowded together in the apical turns, whereas more space is allotted in the basilar turns to the upper octaves.) We are therefore justified in accepting a "place" theory of pitch appreciation. There is a formidable array of evidence for the belief that the localizing mechanism comprises the fibers of the basilar membrane acting as a graduated system of resonators. The projection of the organ of Corti to the primary auditory center and to the cerebral cortex (p 1044) with the preservation of a spacial arrangement also speak for a place theory.

Observations and experimental results bearing upon the questions discussed in this section will be briefly cited.

(1) *The structure of the cochlea* It is scarcely reasonable to regard the design of the cochlea, which upon close examination appears to be so admirably adapted to the selective reception of vibration frequencies, as being quite without meaning. Surely a much simpler mechanism would suffice for the stimulation of the nerve endings were the analysis of sound undertaken by the cerebral centers.

(2) Birds and certain other animals whose cries have a narrow range of pitch have short basilar membranes with little variation in the lengths of their constituent fibers.

(3) A change in the form of a sound wave caused by simply altering the phases of the combined waves in relation to one another does not alter the pitch or quality of the sound. This fact alone is strongly in favor of the resonance theory and the results of an ingenious experiment performed by Hartridge lend further support. The phase of a tone sounded by a siren was altered by half a wave length (180°). The change in phase caused a brief period of silence (the "phase change beat"), the tone then returned and increased rapidly to its original intensity, there was no change in pitch. This phenomenon is taken as indicating the existence of resonators in the cochlea, being compared to the effect caused by a troop of men marching in step across a suspension bridge and

then suddenly and in unison changing step. When the change in step is made, the bridge, which had been vibrating "in tune" with the tramping feet, is brought to rest for an instant, since the steps are now out of phase with its vibrations. But after this momentary period of "silence" the structure again picks up the rhythm, its motions soon reaching their previous amplitude.

(4) *The correlation of lesions localized to definite cochlear levels with loss of hearing over limited ranges of pitch* (a) Boilermakers and others who work amidst loud, clanging sounds sometimes become deaf to tones of the same pitch as that of the noise, though retaining their hearing over other parts of the scale. Degenerative changes are found in that part of the basilar membrane which in accordance with the "place" theory corresponds to the range of the deafness. (b) Crowe and his associates examined *post mortem* the internal ears of 79 subjects of high tone deafness. Three-quarters of these showed atrophy of the organ of Corti in the basal turn, or of the nerves supplying this part of the cochlea. These changes were much more extensive and severe than any found in the ears of 200 subjects with normal hearing. (c) "Islands" of hearing can sometimes be demonstrated in persons congenitally deaf to all other tones. It is impossible to account for such a phenomenon by any theory which attributes sound analysis to the cerebral cortex. It is inconceivable that the brain can perceive one set of impulse frequencies but fails in respect to all others, or that the basilar membrane vibrates as a whole to certain small ranges of vibration frequencies, as the telephone theory demands, but is unresponsive to those higher or lower in the scale. (d) The experiments of Witmach and later of Yoshii and other investigators afford important evidence of some mechanism whereby sound vibrations are localized upon the basilar membrane according to their frequencies. Guinea-pigs were exposed over long periods to loud sounds of constant pitch. When the stimulating tone was of high frequency histological examination of the cochlea after death revealed degenerative changes in the organ of Corti of the basal turn, tones of medium pitch resulted in similar changes in the middle turn, structural changes at the apex were not produced, however, by tones of low pitch. Pavlov has shown, however, by the conditioned reflex method that destruction of the upper part of the cochlea in the dog causes deafness to frequencies below 600 c.p.s. (see footnote, p 1065). (e) Davis and Derbyshire and associates have recently investigated the effects of prolonged stimulation upon the labyrinth of guinea-pigs and upon the threshold of the electrical response of the cochlea. It was found that exposure for 75 days to frequencies of from 600 to 800 c.p.s. and intensities of from 60 to 95 decibels occasionally produced slight wide-spread cochlear damage and a correspond-

ing rise in the threshold of the cochlear response. However, exposure to tones of 2500 c.p.s. and intensities of about 100 decibels for 40 days or so caused more or less extensive damage of the external hair cells of the organ of Corti in the middle part of the second turn of the cochlea and a corresponding rise in the threshold of the cochlear response, the highest threshold being for tones of 1200 c.p.s. In previous experiments it was shown by these observers that the cochlear response was a reliable index of auditory function, for its threshold was found to agree with that as determined by the method of conditioned reflexes. (f) Stevens, Davis and Lane recorded the electrical response before and after drilling through the cochlear wall at different levels and damaging the organ of Corti. Good correlations were obtained between the levels of the lesions and the changes in threshold of the cochlear responses at corresponding frequencies. The diagram shown in fig 78.8 was drawn from the results of these experiments.

THE BASIS FOR THE PERCEPTION OF LOUDNESS
We have seen in considering the sense organs that

the frequency of the impulses discharged along the individual nerve fibers, and probably also in some instances at least the number of functioning fibers, are the factors determining the intensity of the resulting sensation (p 946). The loudness of a sound is now also believed to be dependent upon the frequency of the impulses in the individual nerve fiber and, but to a more important extent, upon the total number of active fibers. With a sound of low intensity a narrow section of the basilar membrane tuned to a particular wave frequency is set into vibration and relatively few fibers are stimulated, whereas a loud sound sets in motion the basilar membrane on either side of this region of maximum vibration and a much larger number of fibers is stimulated. Thus, to quote Hallervell Davis "—pitch is a function of *where*, while intensity is a function both of *how much* of the basilar membrane is disturbed and of an increase in the rate of discharge of many of the activated fibers."

THE EUSTACHIAN TUBE DEAFNESS HEARING TESTS THE HARMFUL EFFECTS OF NOISE

THE FUNCTION OF THE EUSTACHIAN TUBE
THE EFFECTS OF OCCLUSION The Eustachian tube affords the only means whereby the pressure of air within the middle ear can be equalized with that of the atmosphere. The pharyngeal orifice of the tube is closed at ordinary times, but opens during swallowing, yawning or when a high pressure is created in the nasopharynx, as by blowing the nose or making a forced expiration with the nostrils and mouth closed (Valsalva's experiment). During swallowing the orifice is opened by the contraction of small muscles—the *salpingopharyngeus* and *dilatator tubae*—attached to its margins. The pressure within the tympanum may be raised above atmospheric pressure by the accumulation of inflammatory exudate, or become subatmospheric as a result of occlusion of the Eustachian tube. In the latter event air is gradually absorbed from the middle ear into the blood stream (see p 432) and the partial vacuum thus created causes retraction of the drum membrane. The inward displacement of the membrane may be so great as to bring it into contact with the inner wall (promontory) of the tympanum, the cavity being almost obliterated. Deafness results, which is relieved by catheterization of the Eustachian tube or by inflating the ear by means of a Politzer bag, and thus equalizing the pressure on the two sides of the membrane. Temporary occlusion of the tube caused by swelling of the mucous lining is not an uncommon occurrence during an ordinary cold. A sensation of fullness in the ears is experienced, and usually some loss of hearing which is quickly relieved, as a rule, by swallowing or by blowing the nose and thus forcing air up the Eustachian tube. The latter practice is decidedly risky since it is likely to force infective material into the ear and cause acute inflammation.

The tympanic membrane is very sensitive to any difference in pressure on its two surfaces, and during rapid changes in altitude, as in aeroplane ascents and descents, annoying aural effects may be produced. In experiments reported by Armstrong and Heim, bulging of the membrane accompanied by a sensation of fullness occurred when the atmospheric pressure was reduced by an amount corresponding to an altitude of

from 100 to 180 feet. When the reduction in pressure amounted to 15 mm Hg, which corresponds to an elevation of 500 feet above sea level, a distinct "click" was felt in the ear as the drum membrane snapped back into its normal position. These effects were due to the higher pressure in the tympanum forcing the Eustachian tube and thus equalizing the pressure on the two sides of the membrane. With further progressive reduction in atmospheric pressure a similar sequence of events was observed, except that the successive "clicks" occurred with a reduction in outside pressure of only 11.4 mm Hg. This is taken to indicate that the pressures within and without the tympanum are not fully equalized at these altitudes, but that, though the Eustachian tube is forced open at an excess pressure of 15 mm Hg in the middle ear, closure of the tube occurs while the inside pressure is still 3.6 mm Hg above that of the atmosphere.

The results of raising the outside pressure above that in the middle ear, as in aeroplane descents, are quite different. The Eustachian tube acts like a valve, remaining firmly closed against any degree of external pressure. Up to a pressure of from 80 to 90 mm Hg, the tube can be opened by swallowing, but at pressures greater than this the walls of the lower part of the tube are collapsed and held so firmly in this position by the partial vacuum in the middle ear that the dilator muscles are powerless to open it. Rupture of the drum membrane which, according to these observers, is accompanied by severe piercing pain, a loud explosive report in the ear and a sensation of being struck on the side of the head, occurs when the difference in atmospheric pressure and the pressure in the middle ear is from 100 to 500 mm Hg. Nausea, vertigo and general shock are experienced. A pressure difference great enough to rupture the drum membrane in a conscious person is unlikely to occur unless the descent is very rapid (4000 feet per minute) because, even though there is no sensation to give warning of the increased pressure upon the ear drums, most persons swallow automatically every minute or so. Rupture of the membrane might easily occur, however, during sleep, in an unconscious patient in a hospital plane, or in a person with an occluded Eustachian tube.

DEAFNESS Defective hearing may result from disease or abnormality affecting any part of the auditory mechanism. The transmission of sound to the internal ear may be interfered with as a result of some obstruction in the external auditory

meatus, of failure of the Eustachian tube to maintain communication between the pharynx and the middle ear, or of some disease or defect of the middle ear itself. Loss of hearing resulting from any of these causes is termed *transmission* or *conductive deafness*. Deafness due to loss of function from whatever cause of the receptor organs of the internal ear or of the auditory nerve is termed *perceptive deafness*, and that resulting from a lesion of the auditory pathways or of the auditory center is called *central deafness*.

Transmission deafness Wax or a foreign body in the external auditory meatus or some developmental abnormality of the external ear may, by shielding the ear drum from the sound, cause deafness which, of course, varies in severity with the degree to which the meatus is obstructed. Deafness, usually of a temporary nature, commonly results from acute inflammation of the middle ear—*acute otitis media*—which is caused in the great majority of instances by the passage of infective material from the nasopharynx along the Eustachian tube during an ordinary cold, influenza or sore throat. During the acute stage of *otitis media* the ear becomes filled with fluid and is very painful, the accumulation of fluid raises the pressure within the tympanum and bulges the ear drum outwards. The inflammation may progress to pus formation which, if the condition is left to itself, is followed by perforation or rupture of the drum membrane. The surgeon endeavors to prevent rupture by making an incision at a dependent point so as to provide free drainage. The incision soon heals, leaving, as a rule, little or no loss of hearing.

Chronic or subacute *otitis media* causes progressive impairment of hearing due to fibrous adhesions which limit the movements of the ossicles, ankylosis of the incudomalleolar joint, or fixation of the base of the stapes in the oval window. Other factors which may be responsible for the loss of hearing are thickening of the drum membrane and occlusion of the Eustachian tube. In chronic middle ear disease the subject as a rule suffers a greater degree of hearing loss in the lower and middle tones of the scale, in *perceptive deafness* hearing is impaired mainly for the higher frequencies.

Otosclerosis is a chronic ear affection associated with progressive loss of hearing, especially for low tones. The deafness is due to osseous changes which commence in the outer wall of the labyrinth, particularly around the margins of the oval window. Limitation of movement of the base of the stapes results, the ossicle ultimately becoming fixed in the oval window. The pathological process commences with absorption of bone, this is followed by osseous overgrowth, the new formation having the structure of osteoid tissue containing wide spaces filled with connective tissue, rather

than of true bone. Osteoblasts are numerous. The disease extends inwards, involving the bony labyrinth, degeneration of the hair cells of the organ of Corti is found in most instances. Several theories have been proposed to account for otosclerosis, namely, that it is due to a disturbance of calcium metabolism associated with parathyroid dysfunction, that it is caused by infection following inflammation of the middle ear, or that it is a reaction set up by diminished blood supply to the osseous labyrinth caused by vascular sclerosis. Little evidence can be cited in support of any of these theories. The disease shows a familial tendency.

When the footplate of the stapes is rigidly fixed in the oval window the function of transmitting air-borne vibrations to the perilymph devolves upon the round window. The membrane here must then vibrate in sections which move in opposite phases—as one part of the membrane moves in the other moves out. The amplitude of the movements is minimal. For the reason mentioned on page 1200 no appreciable movement of the membrane as a whole is possible. An operation devised by Holmgren and modified by Sourdille, by Lempert and by others, with the view of increasing the amplitude of vibration of the basilar membrane in otosclerosis has so far yielded most encouraging results. The surgical procedure consists in drilling through the osseous wall of a horizontal semicircular canal and covering the gap with a cutaneous-membranous flap reflected from the depth of the external auditory canal and the adjacent portion of the tympanic membrane. By thus providing an artificial second window, pressure variations incident upon the round window are readily transmitted through the internal ear.

Perceptive deafness may result from injury which causes detachment of the sensory cells of the organ of Corti, from inflammatory conditions of the labyrinth, from degeneration of the sensory cells following prolonged and strong stimulation (p. 1207) or from injury or disease of the auditory nerve, such as fracture of the base of the skull or tumor in the region of the cerebello-pontine angle. A tumor in the latter situation is likely to involve the vestibular division of the nerve as well. Deafness is not infrequently due to a developmental defect of the internal ear, the degree of deafness varying in different instances from a slight impairment to deaf-mutism. The latter condition, though frequently congenital, may result from some disease of the ear during infancy which leads to complete loss of hearing. In deaf mutes the semicircular canals are not infrequently undeveloped as well as the cochlear part of the labyrinth, in such persons the responses to the ordinary tests of vestibular function (e.g., rotation, p. 979) are absent.

Central deafness A unilateral lesion of the lower part of the auditory pathway, e.g., in the medulla, before any of the fibers of the cochlear division have

crossed to the opposite side of the brain stem causes loss of hearing in the ear on the side of the lesion. But some fibers ascend uncrossed, thus each ear is bilaterally represented in the cerebral cortex. A unilateral lesion of the auditory pathways anywhere above their decussation therefore causes little or no impairment of hearing, and any which may result affects both ears. Slight if any loss of hearing follows the complete surgical removal of one temporal lobe.

Paracusis Willisii is the term given to the curious phenomenon first described by Willis (1630), that a deaf person hears better in noisy surroundings, as in traffic, trains, etc. There are two main forms—false and true. *False paracusis* is common and easily explained upon the basis of masking (p. 1196). *True paracusis* is rare, there is a real increase in auditory acuity, for it is not only conversation which is heard better but a sound kept at a constant intensity is also rendered more distinct by the presence of noise. The phenomenon has not received a satisfactory explanation. Some believe that it occurs only when there is some stiffness in the articulations of the ossicles. The noise vibrations, it is thought, by “shaking up” the joints render the ossicular chain more efficient in transmitting other weaker sound waves. Other suggestions are that the noise increases the blood supply to the cochlea, or that through the jarring effect of the vibrations the auditory nerve is rendered more irritable.

HEARING TESTS

Disease confined to the middle ear affects the transmission of air-conducted sounds only, conduction through the bones of the skull is not interfered with. When, for example, a vibrating tuning-fork is held to the ear of a subject with middle ear disease, he may fail to hear it, but when the shaft of the fork is pressed against his skull (e.g., the mastoid process) the sound is heard at least as well by him as by one with normal hearing. In deafness due to disease of the internal ear or of the auditory pathway the perception of bone-conducted sounds as well as those transmitted through the middle ear is impaired.

The following tests are employed to distinguish perceptive from transmission deafness.

Weber's test The tone of a tuning-fork applied to the forehead of a person with normal hearing is referred to the mid-line. In unilateral middle ear deafness the sound is localized in the diseased ear, whereas in deafness due to disease of the labyrinth or of the auditory nerve of one side it is heard best or only in the normal ear. This apparently arbitrary result is to be explained upon the basis of masking. In unilateral middle ear deafness, the cochlea of the affected side is shielded from air-borne sounds as a result of defective transmission, whereas on the normal side the tone is masked,

it follows that in a perfectly silent room the sound would not be localized to either side. A person with normal hearing ordinarily hears more acutely in one ear if the opposite ear is closed by the finger, but in perfectly silent surroundings the hearing in one ear is not improved by blocking the other. *Rinne's test* Each ear is tested separately, the auditory meatus of the opposite side being blocked by the observer's finger. A vibrating tuning-fork is applied to the mastoid process, when the sound is no longer audible to the patient the fork is brought close to the corresponding auditory meatus. The subject of middle ear deafness does not hear the sound after it has ceased to be heard through the bone, the test is then said to be negative (*Rinne negative*). A person with a normal middle ear hears the sound for a short time through the air after it is inaudible through bone (*Rinne positive*). *Schwabach's test* consists of the accurate measurement in seconds of the time interval during which the sound is heard through bone (bone-conduction test) or through the air (air-conduction test). Or the length of the interval in each instance may be compared with that during which it is heard by the examiner. For example, at the moment that the patient no longer hears the sound through bone, the fork is placed upon the mastoid process of the examiner, or after the patient ceases to hear the fork through air it is brought to the examiner's auditory meatus. In both forms of the test the length of time during which the examiner hears the sound after it is inaudible to the patient gives a measure of the hearing loss. A patient with middle ear deafness hears the sound by bone-conduction for a longer time than the examiner, in whose ear the tone of the fork is masked by room noise. In a perfectly silent room, of course, the patient's hearing through bone would be no better than the examiner's. In perceptive deafness the tone is heard through bone for a shorter time than the normal.

TESTS OF AUDITORY ACUITY The foregoing tests are employed in determining the part of the auditory apparatus affected. Other test sounds, such as the observer's voice (whisper), the click of a coin or the tick of a watch, are used in determining the threshold of hearing. Since a sound varies inversely as the square of the distance from its source, these tests have no meaning unless the maximum distance at which the sound can be heard by the subject is known and compared with that which can be heard by the average normal ear. The subject's hearing is then expressed as a fraction of the normal—the maximum distance at which the sound is just audible to the patient over that at which it is heard by a person with average hearing under the same conditions. In quiet surroundings a good watch is heard by the normal ear at a distance of about 3 feet and words spoken in an ordinary whisper at about 30 feet. These tests are admittedly rough and give little information of the threshold of hearing at different pitches. In such a determination a series

of tuning-forks may be used. The test fork is given a "standard" blow and then held as close to the ear as possible with the flat surface facing the meatus. The time, t , in seconds from the moment that the blow is struck until the sound becomes inaudible to the patient, is compared with the normal time, t_0 . The degree of hearing loss is then derived from the difference, $t_0 - t$.

The audiometer. Within recent years an instrument known as the audiometer has been introduced for the rapid and more precise testing of auditory function. The apparatus is designed upon principles similar to those used in radio sets, the test tones being generated electrically and conveyed by wires to a receiver

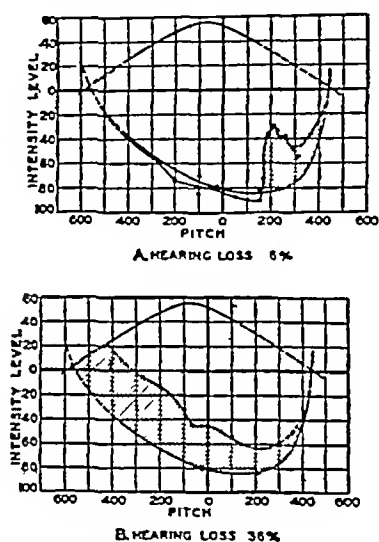


FIG 791

applied to the patient's ear. The tone is varied in intensity or pitch as required by means of dials upon the front of the instrument. The intensity (in decibels above the audible threshold) or the pitch is indicated directly on the respective dial. In one form of the instrument compound tones covering the range of speech frequencies are employed, another type replaces the tuning-fork test, a series of pure tones with frequencies ranging from 64 c.p.s. to 8000 c.p.s. being employed. In a third type speech sounds from a gramophone record are led through the instrument and then to a receiver applied to the patient's ear. The results of the test are plotted as shown in fig 7514. Such a record is called an *audiogram*. Special receivers have also been designed for bone conduction tests. Alternating electrical currents are transformed into mechanical vibrations applied to the skull.

THE HARMFUL EFFECTS OF NOISE Noise is recognized as an annoyance and a general nuisance but is not, as a rule, classed as an influence detrimental to health or efficiency. Yet there is no question that, quite apart from the injurious effects of prolonged and loud sounds upon the internal ear (p 1207) or of loud reports upon the ear drum and mechanism of the middle ear, noisy surroundings can lead to mental irritability, "nervous strain" and increased expenditure of energy. Noise reduces efficiency and accuracy in the performance of manual work, e.g., mail sorting, typesetting, operating factory machines, etc., which demand some degree of attention. The distracting effect of noise upon mental tasks and its interference with restful sleep are too obvious to require comment. That it has a deleterious influence upon persons seriously ill is unquestionable, it is therefore reasonable to believe that, though a healthy person appears to endure almost any degree of noise without immediate detriment to his well-being, he is not immune to its ill effects. There are many observations confirming this. Watkin-Thomas and Yates cite the case of an elderly office worker who was moved from a quiet to a noisy environment. Although his disposition had previously been kindly and courteous it became altered in an unaccountable way, not until he took up his quarters again in a quiet room did his good nature return. In factory workers the accuracy of certain types of operation may be reduced by from 50 to 60 per cent by noise. In a telephone exchange wrong numbers were observed to decrease by 42 per cent when the noise was lowered from 5 to 3.5 bels. It has also been shown that noise increases the oxygen consumption for a given piece of light work, e.g., typing, by from 20 to 25 per cent. The extra energy expenditure is due apparently to increased action of the heart, more rapid respiratory rate and greater muscular tone. In animals, e.g., cats, dogs, rabbits and frogs, such reactions, as well as a rise of blood pressure, can be shown to result from continuous noise or from a sudden loud sound. Beatty points out that there are only three things which evoke a fear reaction in very young infants, namely, letting them fall even a short distance through the air, holding their arms to their sides, or a sudden loud noise.

CHAPTER 80

THE CHEMICAL SENSES

TASTE, GUSTATION

Taste and smell are chemical senses, that is to say, the receptors (chemoreceptors) for these senses respond adequately to chemical stimuli. In order, therefore, for a substance to arouse a sensation of taste it must be dissolved—either taken in solution or dissolved in the saliva, a solid taken into a perfectly dry mouth is tasteless. For this reason the organs of taste or *taste buds* are present only upon a moist surface, being confined to the mouth region of all air-breathing vertebrates, but may be anywhere upon the body surface of aquatic forms.

THE ORGANS OF TASTE The taste buds of man are mainly situated on the tongue but a few are also found in the mucous membrane covering the soft palate, fauces, epiglottis and in the region of the arytenoid cartilages. Taste buds are more widely distributed in children, and are especially plentiful over the anterior part of the tongue. In the adult they are much fewer at the tip of the tongue and are almost absent from the middle third. In most fishes the skin of the general body surface is plentifully supplied with taste receptors, and in the catfish and certain other species of fish they are contained in the filiform processes known as barbules projecting from the snout and angles of the mouth. In insects (flies, bees) taste receptors are located at the end of the proboscis antennae or upon the tarsal segments of the legs.

The mucosa of the human tongue is studded with large numbers of small elevations—the *lingual papillae*—caused by projections of the corium. The papillae are of three main types, filiform, fungiform and vallate. The *filiform papillae* are very minute conical structures covering the anterior two-thirds or so of the dorsal surface of the tongue. They are arranged in rows running roughly parallel with the rows of vallate papillae. The *fungiform papillae* are considerably larger than the preceding type, round in shape and situated mainly at the tip and edges of the tongue. The *vallate papillae* are much larger and become especially prominent posteriorly where from six to twelve are arranged conspicuously in the form of a V with its limbs open anteriorly. A vallate papilla consists of a central round elevation with perpendicular sides and surrounded by a sulcus, the taste buds are situated in the mucosa forming the walls of this circular trench. The filiform papillae rarely contain taste buds, but each fungiform papilla usually holds from 8 to 10 embedded in the epithelium covering its free surface.

A section of a taste bud is shown in figure 80 1. It

measures about 70μ long and 50μ broad, and lies with its long axis perpendicular to the epithelial surface. It consists of groups of *supporting cells* (*peripheral supporting cells*) shaped somewhat like the sections of a musk melon and arranged side by side to enclose a small oval chamber which opens superficially through a circular gap—the *inner taste pore*—surrounded by the converging ends of the supporting cells. The inner taste pore usually leads into a short canal which opens in turn through the *outer taste pore* upon the surface of the tongue. The cavity of the taste bud is occupied by other supporting cells (*central supporting cells*) in the intervals between which the taste receptors (*taste cells*) are lodged. The taste cell is spindle-shaped and provided with a fine hair-like process which projects through the inner taste pore into the short canal mentioned above. The taste bud contains a variable number of these sensory cells, usually from 5 to 18. Nerve

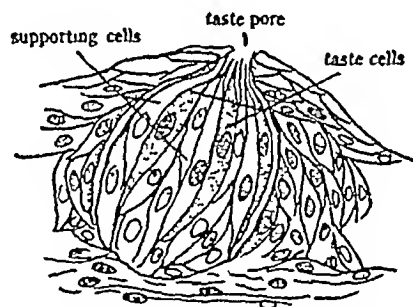


FIG 80 1 Vertical section through a taste bud

fibers after losing their medullary sheaths penetrate the bud and arborize upon the surface of the taste cells.

The chief nerves of taste are the *chorda tympani* branch of the facial nerve and the *glossopharyngeal nerve* (see fig 80 2). The former nerve supplies the taste buds over the anterior two-thirds of the tongue, the latter is distributed to the posterior third. The *vagus nerve* innervates the few taste buds which are present in the region of the epiglottis and arytenoid cartilages. The *trigeminal nerve* mediates common chemical sense (p 1224) and sensations of touch, temperature and pressure (common sensibility) from the entire buccal mucosa, it does not contain taste fibers. Cushing observed, for example, that removal of the semilunar ganglion did not cause any permanent loss of taste. Section of the nerves of taste in animals is followed by degeneration and gradual disappearance of the taste buds. Olmsted has shown in experiments upon the catfish that taste buds reappear upon regeneration of the nerve fibers. The latter evidently exert a

formative influence, possibly through the medium of a chemical substance, upon the development of the taste organs

The central connections of the nerves of taste are described in chapters 66 and 68 (see fig 80.2)

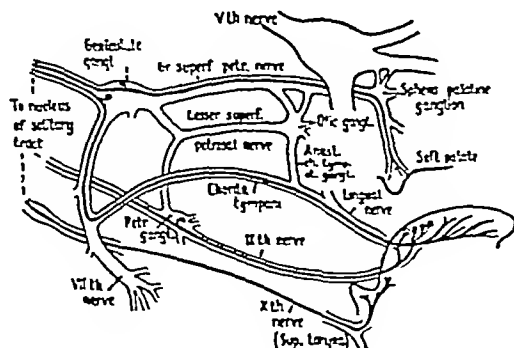


FIG 80.2 Diagram of the course of the gustatory fibers (Modified from Brodal, after Rowbotham)

THE SENSATIONS OF TASTE

There are four *simple, primary or fundamental* tastes—*sweet, sour (acid), salty and bitter*. Two others are sometimes added, namely, *alkaline* and *metallic*¹. The various other tastes which we experience are (a) blends of two or more of the primary sensations or (b) combinations of the latter with sensations aroused by the stimulation of the nerves of common sensibility. For example, ginger is recognized not only by its taste (i.e., through impulses from the taste buds), but also by the burning sensation caused by the excitation of the ordinary sensory nerves of the mouth and also, we may add, by its odor. Many other substances, such as fats and oils and pungent condiments, are “felt” as well as tasted.

Many of the finer flavors are in reality sensations of smell, and olfaction enters very largely into many of the sensations which we generally class as tastes. For this reason when the nose is held or the nasal passages blocked, as during an ordinary cold, our sense of taste seems blunted. It may then be impossible if two bland foods are

¹ Opinions differ as to the nature of these two sensations, most investigators contending that the former is a compound sensation, resulting from the excitation of several types of end organs, including those for sweetness and for touch. Similarly, the metallic taste caused by the salts of heavy metals, copper, silver, mercury, etc., is believed to be a complex of sour and sweet. Some maintain indeed that it is due chiefly to the stimulation of olfactory receptors.

of the same consistency to distinguish between them, thus an apple and a pear, or a turnip and a potato, taste alike. On the other hand, certain substances which we think that we detect by smell are actually tasted. The sweetish smell of chloroform is an example, the vapor reaches the taste buds in the inspired air.

The four primary gustatory sensations are not aroused with equal intensity over all parts of the tongue. Apparently there is a functionally distinct type of receptor for each primary taste, and the distribution of each type is not uniform over the lingual mucosa. End organs sensitive to sweet and salty materials are most plentiful at the tip, those responsive to acid are distributed mainly along the margins, while those aroused by bitter substances are towards the base of the tongue and in the region of the epiglottis. These facts are recognized generally in practice, for one would no

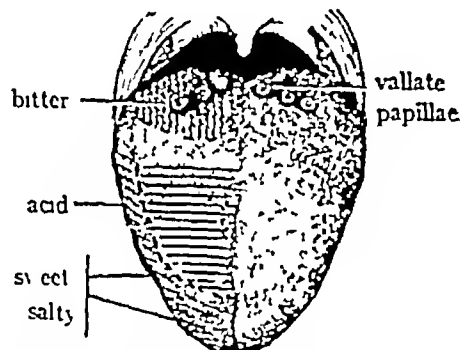


FIG 80.3 Showing distribution of primary taste sensations on one side of the tongue.

more think of sipping beer than he would of gulping a glass of wine, and a child prefers to lick rather than munch a stick of candy. The taste receptors adapt rather rapidly, for this reason food which we enjoy is moved over the surfaces of the tongue and mouth and brought into contact continually with fresh receptors. Some substances stimulate two types of taste bud. For example, *sodium salicylate*, *flamivose* and *parabrom benzoic-sulphinide*, a substance related to saccharine, give a sweet taste when applied to the tip of the tongue but when swallowed, and thus brought into contact with the vallate papillae, taste bitter. *Ortho-benzyl-benzoate*, on the other hand, gives a bitter followed by a sweet taste. *Magnesium* and *sodium sulphates* are salty-bitter, causing a salty taste at the tip of the tongue and a bitter taste at the base. When the papillae are explored individually with

different sapid substances it is found that the filiform type are insensitive. Of the fungiform papillae, some respond to both sweet and salty compounds, others to acid and sweet and others again to bitter and acid. A few respond to all four types of stimulus. These results indicate the existence of functionally distinct types of taste receptors and that different types are present in the same fungiform papilla. Only taste buds responsive to bitter substances are present in the vallate papillae (fig. 80.3).

The sense of taste may be aroused by substances reaching the taste buds in the blood stream. Thus the intravenous injection of histamine causes a metallic taste, glucinum a sweet taste, and in jaundice a bitter taste may be experienced as a result of the high concentrations of biliary constituents in the blood.

Theories of taste perception. The results of experiments in which action potentials were recorded from single nerve fibers when sapid substances were applied to the tongue strongly support the conception of four specific types of taste receptor. Pfaffman identified three types of nerve fiber in the cat: (a) those which responded to acid, (b) those which responded to acid and salt, and (c) those which responded to acid and quinine. Though fibers responsive to sweetness were not found by this investigator, others using an improved method have obtained responses with sucrose. No difference was observed in the character of the impulses which might serve to distinguish the type of stimulus employed. Such a finding is in accord with what we know of other sense organs, namely, that the quality of sensation is determined by the central connections of the nerve fiber that, is, upon the part of the brain where the impulses set up in the receptor organ by an adequate stimulus, are delivered.

It is quite evident that stimulation of the different types of receptor in suitable proportions accompanied by the excitation of non-gustatory nerves of the lingual and oral mucosa, as well as of the olfactory receptors, could account for the wide range of taste sensations which are experienced.

The mechanism through which the taste receptor is stimulated, and an impulse set up in the nerve fiber is unknown, but a theory has been proposed based upon the inhibition of enzyme systems in the taste cells by the sapid substance. Bourne, for example, found in various mammals a relatively high concentration of

alkaline phosphatase in the epithelium overlying the taste buds, and El Baradi and Bourne have demonstrated that a 0.05 per cent concentration of vanillin strongly inhibits the action of this enzyme. Also a simple esterase is present in the taste buds in fairly high concentration, it is inhibited by quinine but not by salt or sugar.

TASTE SENSATIONS AND CHEMICAL CONSTITUTION. The *sweet* taste is associated predominantly with organic compounds, especially the *sugars* (e.g., sucrose, maltose, glucose, etc.) certain *polysaccharides*, *glycerol* and other *alcohols*, *aldehydes* and *ketones* of the aliphatic series, and *saccharine*, *dulcine* and *chloroform*. But certain inorganic substances, such as *lead acetate* (sugar of lead) and *alkalis* in high dilution, are also sweet to the taste.

Substances which arouse an acid or a salty taste are always electrolytes, but bitter or sweet substances may be either electrolytes or nonelectrolytes.

Several attempts have been made to relate the sweet taste to chemical constitution. Oertly and Myers, for example, from a study of a large number of sweet organic compounds have proposed the theory that every sweet molecule contains two particular types of radical or an atom upon which the sweet taste depends. One of these they call a *glucophore*, the other an *auxogluc*. A glucophore makes a given compound a potential tastestuff, if it is bound to an auxogluc a sweet compound is produced. Some six glucophores and nine auxoglucs have been identified, four of each are listed in the following table.

Glucophores	Auxoglucs
(1) $-\text{CO}-\text{CHOH}-(\text{H})$	(1) H
(2) $\text{CO}_2\text{H}-\text{CHNH}_2-$	(2) CH_2CH_2
(3) $\text{CH}_2\text{OH}-\text{CHOH}-$	(3) CH_2OH
(4) CH_2ONO_2-	(4) $\text{CH}_2\text{OH}-\text{CHOH}$

Thus the hexoses contain the glucophore (1) and the auxogluc (4), glycerol the glucophore (2) and the auxogluc (3), and amino-acetic acid the glucophore (2) and the auxogluc (1).

The *salty* taste is evoked primarily by inorganic compounds, notably the *chlorides* of *sodium*, *potassium*, *magnesium*, *ammonium* and *lithium*, by certain *sulphates*, *bromides* and *iodides* and by *sodium* and *potassium nitrates*. The saline taste of such compounds is attributed to the anions (Cl , Br , I , SO_4 and NO_3), a conclusion arrived at from a comparison of their tastes when in high dilution with that of an equally weak solution of sodium.

acetate. For example, a 0.04 molar solution of NaCl, of KCl or of LiCl has a slightly salty taste, whereas sodium acetate in equal or somewhat lower dilution is tasteless or at least is not salty. Similarly sodium bromide, iodide or nitrate loses its saline taste at a much higher dilution than does the acetate. Of the halogens the dilution at which the salty taste is just perceptible is highest for the chloride, next for the bromide and lowest for the iodide ion. The saline taste is not confined to inorganic compounds. Certain organic compounds, such as the *hydrochlorides of monomethylamine and diethylamine* also possess this property.

The *sour* taste is produced by acids or acid salts. It is generally agreed that the effective agent is the hydrogen ion. This statement would seem to be contradicted by the fact that solutions of certain organic acids, such as acetic, tartaric, citric, etc., are more acid to the taste than a solution of a mineral acid having a considerably greater hydrogen ion concentration. For example, the acid taste of a solution of acetic acid is about equal to one of HCl in a dilution one third as great. Yet as compared with the latter solution, the solution of HCl, since this acid is highly dissociated, is from 4 to 5 times as great. The greater effectiveness of acetic acid for a given H⁺ ion concentration is attributed to its greater power of penetrating the tissues, and therefore to its greater effectiveness in raising the hydrogen ion concentration within the taste buds. The *astringent* taste is attributed to acid in very high dilution, that is, to a greatly attenuated sensation of sourness.

The *bitter* taste, like sweetness, is associated chiefly with organic compounds, especially the *alkaloids* (*quinine, strychnine, morphine, etc.*) and certain *glucosides*. *Picric acid, dextromannose and bile salts* are among the other bitter organic compounds. Of inorganic substances with a bitter taste are *magnesium, ammonium, and calcium salts*. The bitterness of these salts is due to the cation. A slight change in the chemical constitution of a substance often alters its taste from bitter to sweet. Saccharine, for example, is intensely sweet, but some of its derivatives are bitter, dulcin is some 500 times sweeter than cane sugar yet *phenyl-thio-carbamide*, in which one oxygen atom in the dulcin molecule is replaced by a sulphur atom, is bitter to most persons. Phenyl-thio-carbamide is peculiar in that to 3 persons out of 10 it is tasteless. The taste deficiency ("taste blind-

ness") in respect to this substance is hereditary, being transmitted as a Mendelian recessive. Many organic compounds having a bitter taste contain NO₂ groups. If the molecule contains two such groups the compound is usually, though not necessarily, bitter, if three are present it is invariably so.

INADEQUATE STIMULI Of agents other than chemical which are capable of evoking a sensation of taste by far the most effective is the electrical current. Electrical stimulation by means of the constant current, using a pair of electrodes placed upon the tongue causes, upon breaking the current, a metallic taste which persists for a little time. If one electrode is placed in contact with the surface of the tongue and the other upon some indifferent part of the body, a constant current during its passage causes an acid or alkaline taste, depending upon the direction of the current. If the lingual electrode is the anode an acid taste is experienced, whereas if the cathode is the stimulating electrode the taste is alkaline. Two factors, apparently, are concerned in the production of the acid or alkaline taste, namely, direct electrical stimulation of the taste cells and the production of H⁺ and OH⁻ ions at the anode and cathode, respectively, as a result of electrolysis of the buccal fluids. That a gustatory response can be produced by direct electrical stimulation is evident from the fact that it is more readily aroused by a rapidly alternating current, which has no appreciable electrolytic action, than by a direct current. Furthermore, when two persons are connected each to a pole of a battery and the circuit completed by bringing the tips of their tongues together, they experience different taste sensations, one acid the other alkaline. Now the two sets of taste buds must be exposed to the action of the same ions, the only condition of the experiment which is different in respect to the taste organs of the two persons is the direction of the current. The electrical taste evoked by a constant current is a rather complex sensation and cannot be described as purely acid or alkaline in quality. It frequently has a bitter metallic component which, as mentioned above, is usually the only taste caused by a single break shock. Very probably electrolytic products as well as the direct stimulating effect of the current are responsible for evoking the complex response. The gustatory sensation caused by a single shock is apparently due purely to direct electrical stimulation, since a current of such brief duration would not have any electrolytic action.

Thermal and mechanical types of stimulation may arouse faint sensations of taste but, as a rule, are ineffective.

Thresholds of the primary taste sensations Minimum concentrations of the four main groups of

sapid substances which will evoke the corresponding sensations are given in the following table

Sensation and substance	Concentration
Sweet cane sugar	1 part in 200
dulcin	1 part in 100,200
α -amylaldehyde	
perillaldehyde	1 part in 600,000
Salty, sodium chloride	1 part in 400
Acid, hydrochloric	1 part in 15,000
Bitter, quinine	1 part in 2,000,000
strychnine	1 part in 2,500,000

AFTER TASTE AND TASTE CONTRASTS The sense of taste exhibits phenomena analogous to positive after images and successive and simultaneous contrast, which have been described for vision (p 1140) It is a familiar experience that the tastes of certain substances (e g , quinine) "cling" to the tongue But it is unlikely that the persistent taste is a true after sensation comparable with an after image, it is most probably due simply to the continued action of the stimulating agent which, having entered the taste pore, is removed with difficulty by the saliva or even by rinsing the mouth with water On the contrary, the metallic taste which outlasts a single break shock is in all likelihood an example of the persistence of sensation

Several observations exemplifying *successive contrast* can be cited A sweet taste is enhanced by a preceding salt or bitter taste and vice versa In the same way sour and sweet tastes intensify one another Even distilled water tastes sweet after rinsing the mouth with a weak solution of sulphuric acid, and lemon juice seems much more acid following a sweet stimulus Other examples which should probably be placed under the heading of successive contrast are the sweet taste which is experienced upon smoking a cigar or cigarette after washing out the mouth with a weak solution of copper sulphate, and the bitter taste caused by smoking if the tongue or buccal mucosa has been treated with a solution of silver nitrate *Simultaneous contrast* is also demonstrable For example, if one border of the tongue is rubbed with salt the sensitivity of the opposite border to a sweet stimulus is increased This contrast effect must, of course, be of cerebral origin Salt and acid also show simultaneous contrast, but the phenomenon cannot be demonstrated for the bitter taste

The effects of drugs upon taste Certain drugs have a selective action upon the taste sensations, abolishing some while leaving others unaffected For example, after the application of a decoction of the leaves of *Gymnema sylvestre* to the tongue, sweet and bitter

substances cannot be tasted, but saline and acid tastes are retained, and are only slightly if at all depressed *Stovaine* acts similarly to gymnema but is less effective. *Cocaine* abolishes all taste as well as common sensibility, the several sensations disappearing in the following order, pain, bitter, sweet, saline, acid and touch

SMELL, OLFACTION

THE OLFACTORY EPITHELIUM. The mucous membrane lining the greater part of the nasal cavity has no true olfactory function (see p 1219) The olfactory receptors are confined to the nasal mucosa over a relatively small region—the *olfactory area* This area



FIG 804 Upper, human olfactory cleft opened by turning nasal septum (S) upward, black area represents olfactory epithelium (Redrawn from Parker, after Read) Lower, transverse section through human left nasal cavity 1, olfactory cleft, 2, septum, 3, 4 and 5, superior, middle and inferior conchae

comprises, on each side, the walls of a narrow niche (fig 804) formed by the superior nasal concha, the upper part of the septum and the roof of the nose (cribriform plate of the ethmoid bone) The olfactory epithelium differs both in its gross appearance and histologically from the rest of the nasal mucosa It is yellowish or brownish yellow in color, its total area, i e, on both sides of the nose, is about 500 square millimeters²

² The vomero-nasal organ (organ of Jacobson) is a short tubular structure which, though well developed in certain lower vertebrates, is rudimentary in primates It can be identified in a vestigial form in infants but, though it may persist throughout life, it is commonly absent in the adult When present it is situated in the lower anterior part of the nasal septum and opens into the cavity of the nose by a minute pore a

The olfactory epithelium is composed of three types of cell (a) supporting cells, (b) basal cells and (c) bipolar nerve cells. The *supporting cells* are of a very high columnar type with large oval nuclei. Superficially, they form a continuous epithelial surface, except for small round gaps between them through which the olfactory vesicles with their tufts of hairs project (see below). Their cytoplasm contains granules of a golden brown pigment to which the color of the olfactory epithelium is due. Proximally, the supporting cell tapers into a long slender process which extends as far as the lamina propria (fig 80.5). The *basal cells* are squat conical structures which extend for only a short distance above the lamina propria

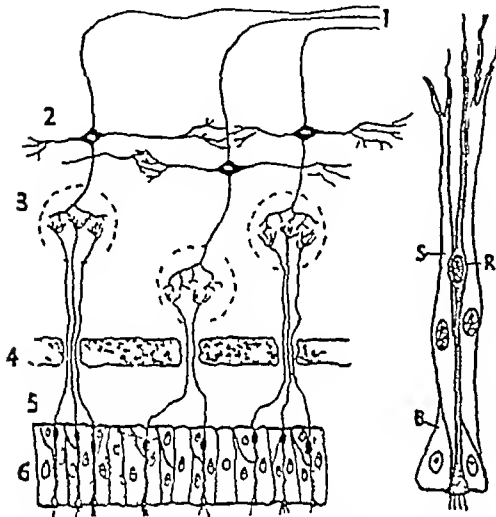


FIG 80.5 Left, diagram showing olfactory epithelium and connections of olfactory nerve fibers 1, olfactory tract, 2, mitral cells, 3, glomeruli, 4, cribriform plate, 5, olfactory nerves, 6, olfactory epithelium. Right, enlarged drawing of cells, B, basal cell, R, olfactory receptor cell, S, supporting cell

They are believed to develop into supporting cells, thus serving as a reserve from which the latter when destroyed can be replaced

The *bipolar nerve cells* are the essential olfactory sense organs. Their two processes arise from opposite poles of the fusiform cell body, the dendrite from its superficial and the axon from its deep aspect. The dendrite is a long straight and relatively stout cylindrical process. It extends to the epithelial surface and projecting through one of the gaps between the sup-

porting cells, expands slightly to form the *olfactory vesicle*. The latter contains from 6 to 8 granules each of which gives rise to a hairlike protoplasmic process. The axon of the bipolar cell proceeds centrally from the deep aspect of the perikaryon and after traversing the lamina propria, joins with the central processes of neighboring cells to form some 20 nerve strands—the *fila olfactoria* or *olfactory nerves*. The latter ascend in grooves in the ethmoid bone and, entering the skull through perforations in the cribriform plate, end within the *olfactory lobe (bulb)* in a tuft of delicate filaments which synapse with dendrites of the *mitral cells* (fig 80.5). The synapses form conspicuous spherical structures called the *olfactory glomeruli*. The axons of the mitral cells constitute the *olfactory tract*. Most of the fibers of the latter are continued into the *lateral olfactory stria* which conveys the impulses to the cortical center for smell (ch 68). The olfactory nerves are non medullated, but possess a neurilemma. The lamina propria of the olfactory mucous membrane contains glands of the tubulo-alveolar type—the *glands of Bowman*. They secrete a serous fluid which bathes the epithelial surface, thus providing a solvent for odorous materials. The fluid is delivered by fine ducts which take a perpendicular course to the surface.

It will at once be recognized from the foregoing description that the end organs of smell differ from those of any other sense in that the cell body of the primary neuron is situated in the peripheral organ itself, and is stimulated directly without the intervention of a specialized receptor cell. No other sensory mechanism possesses both these features. Though the primary neuron of the visual pathway is situated in the retina, the stimulus is received by the rods and cones, pain sensations are subserved by bare nerve endings, the cell bodies of the pain fibers are located in the posterior root ganglia. It is also important to remember that here as nowhere else the nervous system is in direct contact with the external environment. The terminations of the dendrites of the olfactory cells, namely, the olfactory hairs, are covered only by a layer of fluid. Furthermore, the sheaths of the olfactory nerves are continuous with the subarachnoid space. Experimental work indicates that the olfactory nerves constitute one pathway through which the virus of anterior poliomyelitis may reach the central nervous system. Shultz and Gebhardt have shown, for example, that monkeys are protected against the intranasal injection of poliomyelitis virus by a previous section of the olfactory nerves. Intranasal sprays, consisting of solutions of zinc sulphate and other substances have been employed in monkeys with the purpose of blocking these channels. The encouragement derived from the success of these experiments in protecting animals from the disease has led to the trial of similar measures in poliomyelitis epidemics, but unfortunately they do not appear to have any value in reducing the number of cases. It is very difficult,

short distance within the external nares. In the dog and cat it receives both olfactory and trigeminal fibers and contains epithelium similar to that of the olfactory area. The function of the vomero-nasal organ is not known with certainty but its general structure and innervation suggest very strongly that it is a subsidiary olfactory sense organ.

especially in children, to bring the solution into contact with the olfactory epithelium, and this is probably the reason that reliance cannot be placed upon intranasal spraying as a preventive. The recent experiments of C. G. Smith are enlightening. The olfactory areas of rats were treated with a 1 per cent solution of zinc sulphate and examined histologically at periods of from 2 days to 2 months thereafter. In animals so treated, destructive changes amounting even to sloughing of the entire olfactory epithelium were found, and in all cases the bipolar cells showed widespread degeneration. Regeneration of non-sensory cells (supporting and basal) subsequently occurred, but, of course, the nerve cells were not restored. It would appear from these experiments that zinc sulphate solution, if it is to be effective in protecting against the virus of poliomyelitis, must entail permanent loss of the sense of smell.

THE PHYSIOLOGY OF THE SENSE OF SMELL

OLFACTORY SENSATIONS Smell is very closely allied to taste and has been aptly described as "taste at a distance." In many animals the sense of smell is almost incredibly acute, a relatively large part of the brain being given over to it. In the life of such macrosmatic animals the olfactory sense is of paramount importance, warning the animal of the approach of its enemies, guiding it in the quest for food and motivating the sex reflexes. Certain species of moth (e.g., great peacock and banded monk) are credited with a degree of olfactory acuity which seems almost mythical to microsmatic man, the female being able, it is claimed, to attract the male by the odor of its secretions from a distance of a mile or more. The olfactory organs of moths and most other insects are located in the antennae. Even man, in whom smell is a comparatively rudimentary sense, can detect certain substances (e.g., mercaptan and artificial musk) in a dilution of 1 part in several billion parts of air, smell is therefore much more acute than the sense of taste.

The adequate stimulus for the olfactory receptors, as for those of taste, is chemical.³ An odorous material continuously emits particles of molecular size which reach the olfactory area through the air. Substances which pass readily into the gaseous state, such as turpentine, gasoline, the essential oils, etc., have strong odors, whereas non-volatile materials, e.g., the heavy metals, are nearly or

quite inodorous. Arsenic, which ordinarily is odorless gives off a characteristic smell, however, when heated to a temperature at which it volatilizes, and Elsberg, Brewer and Levy found that the olfactory coefficients (p. 1222) of a number of odorous liquids vary directly with their boiling points. The niche at the roof of the nose which is lined by the olfactory epithelium, constitutes a blind pocket, from which the main air currents caused by the ordinary respiratory movements are excluded. Experiments upon the human cadaver have shown that the respired air does not come into direct contact with the olfactory area. A head was bisected in the median plane and the nasal septum replaced by glass. When smoke was forced back and forth through the nose by means of bellows, neither during the artificial inspiratory nor expiratory movement was the current observed to enter the olfactory region. The air flow takes a curved course, the highest point being about the middle of the nose and below the superior concha. The stream reaches a lower level during expiration than in inspiration. In order to excite the olfactory cells the odorous particles must, therefore, be carried upwards from the respiratory passages either by diffusion or by eddy currents. In the aforementioned experiment eddy currents were observed both during the inspiratory and expiratory movements. In the living subject it is probable that the ascending currents are more pronounced during inspiration, at this time air movements caused by convection are likely to occur, due to the mixing of the cooler incoming stream with the warmer air within the nose. But, however produced, whether by convection or simply as a result of the mechanical mixing of the inspired air with that within the nasal passages, eddy currents constitute the main factor in the stimulation of the olfactory endings, and a sharp inspiration is the most effective means by which such currents are set up. When, for example, we wish to smell some particular scent more acutely we automatically make a sharp inspiration or "sniff."⁴ Diffusion is a relatively slow process and is probably of minor importance in bringing the odorous material to the olfactory endings. Even though the nose is filled with odor laden air we cannot smell while the breath is held. It is during

³ An electric current acts as an inadequate stimulus. When the nose is filled with normal saline and a constant current passed through the solution an odor which is difficult to describe is experienced upon opening or closing the current.

⁴ The mechanism of the "sniff" appears to be a compressing together of the septum and the outer wall of the nose at the front of the respiratory passages so as to divert the inspired air to the olfactory area (Ogle).

expiration that odorous materials liberated from the food as it is masticated and swallowed enter the nose through the posterior nares and ascend to the olfactory area

It will be recalled that the olfactory hairs are immersed in a layer of fluid secreted by Bowman's glands. The odorous particles must therefore enter into solution before they can come into contact with and stimulate the sense organs. This fact emphasizes again the similarity between the senses of taste and smell. It is probable that odorous materials before they can act as stimuli must also be dissolved in the substance of the olfactory hairs themselves. These structures, since they are stained best by osmic acid, are believed to be composed largely of lipid material. One would expect, therefore, that odorous substances must be soluble in oil as well as in water, and that those which are most freely soluble in both media would be the most potent in arousing an olfactory sensation. This supposition is borne out to some extent by experiment. Ethyl and methyl alcohols, for example, which are freely soluble in water but only slightly in oil, have weak odors as compared with butyl alcohol which dissolves very freely in oil and is also soluble in water. *Chlorobenzol*, *brombenzol*, *ether*, *citral* and many other substances with strong odors are soluble in both water and oil. It would appear that high solubility in oil is of more importance for olfactory stimulation than high water solubility, for taking two substances one with a high solubility in oil but sparingly, soluble in water, and the other possessing solubilities of a converse kind, the former has the most powerful odor.

Though it is no longer questioned that the olfactory nerves subserve the sense of smell, some of the earlier investigators (e.g., Magendie) contended that olfaction was a function of the trigeminal nerve. The confusion arose from the fact that certain agents, e.g., *ammonia*, *nitric acid fumes*, *chlorine*, *pepper*, *menthol*, *peppermint* and many others, cause nasal sensations, usually described as pungent, acrid, irritating or cooling. These are not true olfactory sensations but are due to the stimulation of the trigeminal, which is the nerve mediating chemical sense and common sensibility in the respiratory part of the nasal mucosa. It is often very difficult, however, to dissociate these sensations from smell when the two types of ending are stimulated concurrently. A similar confusion arises, as already mentioned in the case of taste.

Strong reflex effects, e.g., sneezing, lacrymation, respiratory inhibition, vasomotor reactions, etc., result from irritation of the trigeminal endings, whereas reflexes initiated from the olfactory receptors are as a rule mild in character, and in man are concerned mainly with salivary and gastric secretion. In animals olfactory reflexes play their most important rôle in the reactions of sex and in self preservation—the search for food and protection from enemies. Olfaction is paramount among the senses in its power to awaken a train of associations in consciousness. Everyone is familiar with the strange reminiscent aura of long past events which is aroused by certain familiar scents (see p. 1045).

THRESHOLD STIMULI Among the most effective olfactory stimuli are *artificial musk*, *mercaptan*, *butyric acid*, *iodoform* and *oil of peppermint*. For example, methyl mercaptan (garlic odor) is perceptible to the average person in a concentration of 1/23,000,000,000 of a milligram per cu. cm. of air. Assuming that 50 cc. of air is required for arousing an olfactory sensation, this would mean that 1/460,000,000 mg. of the substance is an effective stimulus. The sense of smell is therefore many thousand times (about 25,000 times in the case of ethyl alcohol) more acute than the sense of taste. The minimum perceptible concentrations of various odorous substances are given in the following table.

Substance	Mg. per liter of air
Ethyl ether	5.83
Chloroform	3.30
Pyridine	0.032
Oil of peppermint	0.024
Iodoform	0.018
Butyric acid	0.009
Propyl mercaptan	0.006
Artificial musk	0.00004

(from Allison and Katz)

SENSORY ADAPTATION The olfactory receptors adapt fairly rapidly. It is a common experience that a disagreeable odor which when first smelt is almost overpowering soon becomes imperceptible. But although lost for one particular odor the sense of smell is retained for others, the phenomenon therefore is not due to fatigue of the olfactory mechanism,⁵ but is an example of sensory adaptation (p. 944). The rate of adaptation varies for different odors. The receptors become insensitive to oil of orange or to oil of lemon after an exposure of from 2.5 to 11 minutes (average

⁵Though it is often referred to as such

3 minutes), whereas cumarin (0.2 per cent aqueous solution) cannot be smelt for longer than from 1.75 to 2.3 minutes, and adaptation for the odor of benzoin is more rapid than for that of rubber.

Olfactory adaptation commences to develop from the moment that the odor is first smelt, the threshold rising gradually until complete insensitivity to that particular odor is reached. Even a previous period of exposure to a given odor raises the minimum concentration at which it is perceived for a considerable length of time afterwards. Elsberg found, for example, that the olfactory coefficients (p. 1222) for peppermint, camphor and sassafras were increased to double their normal values if the subject had previously been smelling these substances, and the sensitivity of a person who had been for a time in the operating theater to the odor of ether was below normal several hours later.

CHEMICAL CONSTITUTION IN RELATION TO OLFACTORY STIMULATION Generally speaking, the olfactory potency of chemical compounds belonging to an homologous series increases progressively from the lowest members of the series to the highest. The odors of the monatomic alcohols, for example, increase in strength from methyl through ethyl, propyl and butyl to amyl, the relative potencies of methyl and amyl alcohols are as 1 to 10,000. Also, compounds which as a group resemble one another in their chemical and physical properties tend to have odors possessing certain common characteristics. For example, the elements sulphur, selenium and tellurium, which belong to the sixth group in Mendeleff's periodic table, when combined with hydrogen, methyl or ethyl, etc., have strong disagreeable smells. Similarly members of the seventh group, chlorine, bromine and iodine have kindred odors, the odors of chloroform and iodoform, compounds of the first and third elements respectively, are linked together by that of bromoform in which the fragrance of chloroform and the unpleasant odor of iodoform can be detected. Of chemically allied organic substances, ethyl, propyl and butyl acetates have an acetic odor, whereas amyl acetate has not, nevertheless the smell of the lowest of the series is linked with that of the highest through the two intermediate compounds. Thus—

Ethyl acetate, acetic odor

Propyl acetate, acetic odor with slight pineapple flavor

Butyl acetate, slight acetic odor with pineapple flavor

Amyl acetate, no acetic odor, strong pineapple flavor

Though the foregoing are interesting examples of chemico-olfactory correlation, it is not possible to make anything more than broad generalizations in respect to chemical structure and smell, for compounds which closely resemble one another chemically may have quite different odors and others which show little resemblance in their chemical or physical properties

(e.g., hydrocyanic acid and nitrobenzene, garlic and certain arsenical compounds, and artificial and natural musk) may smell very much alike. An attempt has been made in the case of aromatic compounds to relate odor to a particular radical on the benzene ring. Hydroxyl, aldehyde, ketone, ester, nitro and nitril grouping—the so-called *osmophoric groups*—have been suggested as determining the character of the odor, it is believed, however, that the latter is dependent not so much upon which particular radical is present as upon the position which any one of them occupies in the benzene ring.

A physical property common to many odors is their strong absorption of infrared rays. The significance of this fact, first remarked upon by Faraday, is unknown.

OLFACTOMETRY The most widely known method of investigating the sense of smell is that of Zwaardemaker. His olfactometer consists of two tubes sliding one inside the other, as illustrated in

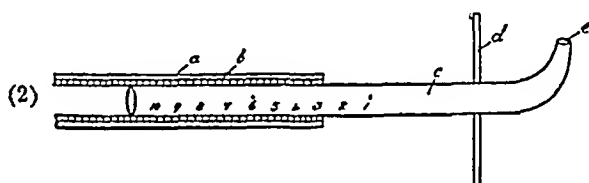


FIG. 80.6 Diagram of Zwaardemaker's olfactometer. See text.

figure 80.6. The inner tube is made of glass and graduated in sections 0.7 cm long. The outer tube, also of glass, has a lining of india rubber, beeswax, sealing wax or some other faintly odorous material. The curved end of the inner tube is introduced into a nostril, the opposite one being closed, the subject breathes quietly. The outer tube is gradually withdrawn, thus exposing a greater area of its inner surface to the air current and thereby increasing the concentration of the odorous particles in the inspired air. The highest figure visible on the inner graduated tube when the odor is just perceived indicates the subject's threshold for smell in units termed *olfactus*. This method gives at the best only approximate results, chiefly because the volume of inspired air drawn through the tubing varies considerably from subject to subject and in the same person at different times or even during a single period of observation. In the *blast method* of Elsberg and Levy this factor is controlled. Thirty cubic centimeters of an odorous liquid (e.g., benzene citral, oil of orange, oil of turpentine, butyric acid, etc.) are placed in the bottle shown in figure 80.7. The right hand tube is

connected to a double nosepiece which fits into the nostrils. By means of a syringe connected to the other tube a measured volume of air is forced from the bottle in one blast at a constant pressure while the subject holds his breath. An equivalent volume of odor-laden air is thus forced into the nose. The volume of the injections is gradually increased in successive blasts until the odor is just perceived and can be named. The smallest volume necessary for identification is called the *minimum identifiable odor* (M I O) or the *olfactory coefficient*.

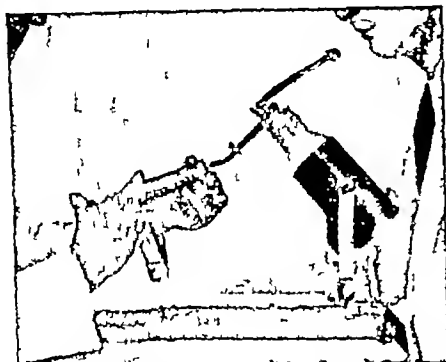


FIG 807 Illustrating the blast method of olfactometry (After Elsberg and Levy)

This method has been employed by Elsberg as an aid in the localization of tumors in the anterior of the skull. In this situation a tumor (e.g., of the frontal lobe) is likely, through direct pressure, to involve the olfactory lobe or tract of one or of both sides, or the olfactory nerves may be torn in fractures through the cribriform plate. Unilateral involvement of the olfactory nerves, lobe or tract raises the M I O or completely abolishes the sense of smell on the affected side. A tumor involving both olfactory lobes or tracts will result in lowered acuity of smell or complete anosmia on both sides. Elsberg states that tests for olfactory "fatigue" give valuable localizing aid, the "fatigue" phenomenon is prolonged beyond the normal limits by tumors within the substance of the temporal lobe, but not by those situated extracerebrally (e.g., beneath the frontal lobe). In cases of a generalized increase of intracranial pressure due to other causes the M I O is often lowered.

CLASSIFICATION OF ODORS The division of odors into categories has proved an extremely difficult problem. There are no basic qualities of olfaction comparable to sweet, salty, sour and bitter tastes. The number of different and distinct smells is legion, and no comprehensive classification upon the basis of chemical constitution or physical

properties can be even attempted. The earliest classification of odors was made by the Swedish botanist Linnaeus (1750). The following one proposed by Zwaardemaker, which is little more than an elaboration of that proposed by Linnaeus, consists of nine categories. It has a purely subjective basis and is therefore of little scientific value.

- 1 *Ethereal odors*, e.g., of fruits, beeswax, ethers
- 2 *Aromatic or resinous odors*, e.g., of camphor, bitter almonds, cloves, lavender
- 3 *Fragrant or balsamic odors*, e.g., of flowers, extracted or artificial perfumes
- 4 *Ambrosial odors*, e.g., of musk, ambergris
- 5 *Garlic odors*, e.g., of garlic, onions and of sulphur and selenium compounds
- 6 *Burning odors*, e.g., of burning feathers, tobacco, roasted coffee and meats
- 7 *Goat odors*, e.g., caproic acid, sweat and ripe cheese
- 8 *Repulsive odors*, e.g., of hyoscyamus and several of the family of the deadly nightshade, bedbug
- 9 *Nauseating odors*, e.g., of excrement, decaying meat and vegetable matter

ACTION POTENTIALS FROM THE OLFACTORY PATHWAYS Gerard and Young have recorded the action currents from the olfactory bulb of the frog. A spontaneous rhythmical discharge occurs in this part of the isolated brain at the rate of about 4 per second. The discharge is taken as representing true automatic activity of the nerve cells of the bulb itself, for the possibility that it was due to the irritation of traumatized structures seems to have been excluded.

In certain fishes (catfish, carp and tench) the olfactory bulb is connected with the forebrain by a nerve strand (the *olfactory stalk*) which measures about 2 cm long and is composed of from 500 to 1000 medullated fibers. Adrian and Ludwig studied the action potentials in the olfactory stalk of the catfish during stimulation of the olfactory end organs. The latter are contained in a small sac which opens through a nipple upon the surface of the skin above the mouth. Potential changes of small amplitude pass up the stalk as long as the preparation survives, though nothing but distilled water has been introduced into the sac. This *resting discharge* is of very low frequency. When the sac was irrigated with fluid containing small fragments of some odorous material, e.g., putrefying earthworms, a burst of impulses occurred at high frequency after a latent period of from 0.5 to 5 seconds or longer. It was found that the irrigation was more powerfully stimulating if the fluid contained small fragments of the material than if it had been filtered.

THE QUESTION OF OLFACTORY DISCRIMINATION The great multitude of distinguishable odors brings

up a question for which no answer is forthcoming, namely, "What is the mechanism underlying olfactory differentiation?" "How do we detect the difference between two scents such as those of the violet and the rose?" Or, to make the question still more difficult, "How does a dog recognize the smell of his master among the smells of other persons?" Were there a limited number of basic odors, as there are fundamental tastes, the problem would not be difficult, but such are unknown. For vision we can postulate three types of receptor, each responsive to a particular set of wave lengths, and each discharging impulses along a specific nerve fiber to the brain (see doctrine of specific nerve energies (p 946). A place theory (p 1207) is capable of accounting for pitch discrimination, and four functionally distinct types of taste bud, each with its specific nerve fiber, for the perception of the fundamental sensations of taste. But it is not conceivably possible that every one of the immense number of different odors is subserved by a specific type of end organ with its own nerve fiber. Furthermore, new and distinct odors are being created in industry every day.

Theories of the mechanism through which an odorous substance sets up an impulse in an olfactory nerve fiber are of necessity almost wholly speculative. The following catalytic theory of olfaction has been proposed by Kistiakowsky similar to that already described for gustation. He postulates a series of catalytic reactions which may be represented as, $A \rightarrow A' \rightarrow A''$, $B \rightarrow B' \rightarrow B''$, $C \rightarrow C' \rightarrow C''$, etc. Each step is effected by a specific enzyme, and each compound formed, A' , B' , C' , etc., stimulates an olfactory receptor and gives rise to a corresponding distinct "basic" odor. Presumably, the odoriferous substance itself inhibits one or more of the enzyme systems. This would lead to differences in the combination of the stimulating compounds, and even a slight change in their relative concentrations could arouse an olfactory sensation. The difficulty here is in defining a basic odor and having to assume an unreasonable number of functionally different receptor cells, for the nerve impulses themselves possess no differentiating characteristics which could serve the brain as cues. But if in addition, areas with different physical or chemical properties were distributed over the olfactory membrane which permitted a selective penetration or adsorption of the odorous substance to or upon the receptors, some sort of pattern of excitation might

be established which could be interpreted by the cerebral cortex as a smell of a certain kind, or the different areas might possess specific affinities for certain chemical groups. A mechanism of this nature would not require an immense number of different types of receptor.

From what we know of other senses we can presume that the *intensity* of the olfactory sensation is related to the frequency of the impulses discharged to the olfactory center. This is borne out by the experiments of Adrian and Ludwig just described.

THE EFFECT OF ONE ODOR UPON THE PERCEPTION OF ANOTHER. Strong odors tend to mask weaker ones. If two scents are of about equal strength a blend of the two is smelt or both are identified, but if one is considerably stronger than the other it alone, as a rule, is smelt. On the other hand, certain pairs of odors in appropriate relative concentrations are antagonistic, and when the two are sniffed together both are diminished. Iodoform, for example, is antagonized by balsam of Peru, musk by bitter almonds and ammonia by acetic acid. Other pairs of neutralizing odors are cedarwood and rubber, beeswax and balsam of Tolu, benzoin and rubber, and camphor and eau de Cologne. Though the neutralizing effect may in some cases be simply chemical or physical in nature, in others there seems to be a true physiological antagonism, for the phenomenon is observed when mixing is avoided by leading the two odors directly one to each nostril.

ANOMALIES OF OLFACTION. Loss of the sense of smell or *anosmia* is not infrequent as a temporary condition, e.g., as a result of inflammation of the nasal mucosa or of the local application of cocaine or adrenaline. Complete and permanent anosmia is rare in otherwise normal persons and is usually due to absence of the olfactory bulb or olfactory nerves, but bilateral or unilateral olfactory deficiencies are frequently associated with lesions in the region of the olfactory lobes (p 1218). Albinoes are said to be anosmic, which suggests that the pigment in the supporting cells of the olfactory epithelium, which possibly serves an essential function, is lacking. Inability to smell certain odors is not uncommon. Some persons, for example, cannot smell hydrocyanic acid, the odor of mignonette, benzoin, methyl alcohol or vanillin cannot be smelt by others. Even such a strong, disagreeable smell as that of a rotten egg or of feces may not be sensed. Partial anosmia for all scents may result from excessive smoking. *Hyperosmia* is not unusual in hysteria, in certain cerebral diseases, in raised intracranial pressure and during the initial stage of the action of cocaine. Olfactory hallucinations may occur in lesions of the temporal lobe (p 1045). In cerebral tumors, especially of the

REFERENCES

SECTION I

CHAPTER 1

THE PHYSIOLOGICAL PROPERTIES, PHYSICAL CHARACTERS AND COMPOSITION OF THE BLOOD

- BLOOR, W R J Biol Chem , 1918, **36**, 49
 CARREL, A J Am Med Assoc , 1924, **82**, 255 J Exp Med , 1922, **36**, 385
 ELMAN, R and HEIFETZ, C J J Exp Med , 1941, **73**, 417
 KAY, H D Note on the chemical composition of the reticulocyte Ibid , 1930, **11**, 148
 KEYS, A J Phys Chem , 1938, **42**, 11
 KEYS, A. and associates Science, 1950, **112**, 79
 MILLER, L L and associates J Exp Med , 1951, **94**, 431
 PHILLIPS, R A , VAN SLYKE, D D and associates Bull of U S Army Med Service, 1943, **71**, 66
 WHITE, A and DOUGHERTY, T F Endocrin , 1945, **36**, 207 Proc Soc Exp Biol Med , 1944, **56**, 26

Monographs and Reviews

- FOLIN, O Non-protein nitrogen of blood in health and disease Physiol Rev , 1922, **2**, 460
 GUEST, G M and RAPPAPORT, S Organic acid soluble phosphorus compounds of the blood Physiol Rev , 1941, **21**, 410
 HENDERSON, L J Fitness of the environment Macmillan, New York, 1913
 HOWE, P E The function of the plasma proteins Physiol Rev , 1925, **5**, 439
 MACALLUM, A B The paleochemistry of the body fluids and tissues Physiol Rev , 1926, **6**, 316
 MYERS, V C Chemical changes in the blood and their clinical significance Physiol Rev , 1924, **4**, 274
 PETERS, J P and VAN SLYKE, D D Quantitative clinical chemistry Vol I The Williams & Wilkins Co , Baltimore, 1932
 SVENSSON, H. Electrophoresis by the moving boundary method Arkiv för Keml Mineralogi och geolgi

CHAPTER 2

THE RED CELLS OR ERYTHROCYTES

- HARTRIDGE, H J Physiol , 1920, **53**, lxxx
 KNISELY, M H and associates Arch , Surg 1945, **51**, 220, Trans 4th Conf on Blood Clotting and Allied Problems J E Flynn, Ed , J Macy Found 1951

Monographs and Reviews

- BARCROFT, J The respiratory function of the blood Part I. Lessons from high altitudes Cambridge University Press, 1925

- EAMONS, W F J Physiol , 1927, **64**, 215
 HARROP, G A , JR. Polycythemia Medicine, 1928, **7**, 291
 PRICE-JONES, C Blood pictures, an introduction to clinical hematology William Wood & Company, Baltimore, 1933
 SCHAFER, P W Ann Surg , 1945
 WINTROBE, M M The erythrocyte in man Medicine, 1930, **9**, 195

CHAPTER 3

THE BLOOD VOLUME BODY WATER WATER BALANCE

- BARBOUR, H G and HAMILTON, W F Am J Physiol 1925, **73**, 315
 BARCROFT, J, KENNEDY, J A and MASON, M F J Physiol , 1939, **95**, 159
 BYROM, F B Clin Sci , 1934, **1**, 245 Ibid , 1934, **1**, 273
 CROOKE, A C and MORRIS, C J O J Physiol , 1942, **101**, 217
 FOWLER, W M and BARER, A P J Am Med Ass , 1942, **118**, 421
 GAMBLE, J L, ROSS, G S and TISDALL, F F J Biol Chem , 1923, **57**, 633
 GAUNT, R Endoc , 1944, **34**, 400
 GIBSON, J G and EVELYN, K A J Clin Invest , 1938, **17**, 153
 GIBSON, J G and colleagues J Clin Invest , 1946, **25**, 605, 617
 GREGERSEN, M I, GIBSON, J J and STEAD, E A Am J Physiol , 1935, **113**, 54
 HAHN, P F and associates J Exp Med , 1942, **75**, 221
 HALDANE, J S and SMITH, J L J Physiol 1900, **25**, 331
 KEITH, N M , ROWNTREE, L G and GERAGHTY, J T Arch Int Med , 1915, **16**, 547
 KENNEDY, J A and MILLIKAN, G A J Physiol , 1938, **93**, 276
 McMASTER, P D , ROUS, P and HUDAK, S J Exp Med , 1955, **203**
 PAINTER, E E and associates Am J Physiol , 1948, **152**, 66
 ROBINSON, E A and ADOLPH, E F Am J Physiol 1943, **139**, 39
 SCHILLER, R AND DORFMAN, R I Endocrin 1943, **33**, 402
 SHORE, E and associates Science, 1945, **102**, 489
 ZWEIFACH, B W and associates Am J Anat 1944, **75**, 239 Am J Physiol , 1944, **142**, 80 Ibid , 1944, **150**, 239 Ann Surg , 1944, **120**, 232 Surg Gynec Obstet , 1945, **80**, 593, Ann N Y Acad Sc , 1946, **46**, 533, 571 See also CHAMBERS, R Nature, 1948, **162**, 835

Monographs and Reviews

- ADOLPH, E F The metabolism and distribution of water in body and tissues *Physiol. Rev.*, 1933, 13, 336
- ADOLPH, E F and associates *Physiology of man in the desert* Interscience Publishers, N Y, 1947
- BARBOUR, H G The heat-regulating mechanism of the body *Physiol Rev.*, 1921, 1, 295
- BAYLISS, W M *Principles of general physiology* Longmans, London, 1927
- ERLANGER, J Blood volume and its regulation *Physiol Rev.*, 1921, 1, 177
- HILL, L and CAMPBELL, A *Health and environment* Arnold, London, 1925
- LANDIS, E M Capillary pressure and capillary permeability *Physiol Rev.*, 1934, 14, 404
- MARRIOTT, W M Anhydremia *Physiol Rev.*, 1923, 3, 275
- PETERS, J P *Body water* Thomas, Springfield, 1935
- ROWNTREE, L G The water balance of the body *Physiol Rev.*, 1922, 2, 116
- ROWNTREE, L G and BROWN, G E *The volume of the blood and plasma in health and disease* Saunders, Philadelphia, 1929
- STARLING, E H *The fluids of the body* Constable, London, 1909

CHAPTER 4

THE LYMPH AND TISSUE FLUIDS

- BARNES, J M and TRUETA, J *Lancet*, 1941, 1, 623
- CAIN, J C and associates *Surg Gyn & Obst.*, 1947, 85, 559
- DRINKER, C K, FIELD, M E and WARD, H K *J Exp Med.*, 1934, 59, 393
- DRURY, A N and JONES, N W *Heart*, 1927, 14, 55
- HENSCHEL, A. and associates *Am J Physiol.*, 1947, 150, 170
- LANDIS, E M and GIBBON, J H, JR The effects of temperature and of tissue pressure on the movement of fluid through the human capillary wall *J Clin Invest.*, 1933, 12, 105 See also, Landis and associates The movement of fluid through the human capillary wall in relation to venous pressure and to colloid osmotic pressure of the blood *Ibid.*, 1932, 11, 63 The passage of fluid and protein through the human capillary wall during venous congestion *Ibid.*, p 717
- MCMASTER, P D *J Exp Med.*, 1937, 65, 373, *ibid.*, 1947, 86, 293 *Harvey Lectures*, 1941-42, Ser 37, 227
- MARKOWITZ, C and MANN, F C *Am J Physiol.*, 1931, 86, 709
- MERRILL, A J *J Clin Invest.*, 1946, 25, 389, *Proc. 22nd, Sci Assoc., Am Heart J* 1949, p 34
- NEWMAN, E V *Am J Med.*, 1949, 7, 490
- PARSONS, R J and MCMASTER, P D *J Exp Med.*, 1938, 68, 353, 377

- PETERS, J P *N Eng Med J.*, 1948, 239, 353
- RICHARDS, D W *Am J Med.*, 1949, 6, 772
- WARREN, J V and STEAD, E A *Arch Int Med.*, 1944, 73, 138

Monographs and Reviews

- CUNNINGHAM, R S The physiology of the serous membranes *Physiol Rev.*, 1926, 6, 242
- DAVIS, J O and SMITH, J R Pathogenesis of cardiac edema *Am J Med.*, 1947, 3, 704
- DENZ, F A *Hunger Edema Quart J Med.*, 1947, 16, 1
- DRINKER, C K and FIELD, M E *Lymphatics, lymph and tissue fluid* The Williams & Wilkins Co, Baltimore, 1933
- DRINKER, C K and YOFFEY, J M *The lymphatics, lymph and lymphoid tissue* Harvard University Press, 1941
- LANDIS, E M The passage of fluid through the capillary wall *Harvey Lectures*, 1936-7, Ser 32, 70
- MCMASTER, P D Lymphatic participation in cutaneous phenomena *Harvey Lectures*, 1941-42, Ser 37, p 227
- STARLING, E H *The fluids of the body* Constable, London, 1909

CHAPTER 5

TRANSFUSION THE BLOOD GROUPS

- AMBERSON, W R and associates *Science*, 1933, 78, 106 *J Cell Comp Physiol.*, 1934, 5, 359 *Biol Rev.*, 1937, 12, 46
- BELK, W P and BARNES, B C *Am J Med Sci.*, 1941, 201, 838
- COHN, E J and associates *J Am Chem Soc.*, 1940, 62, 3396 *Chem Rev.*, 1941, 28, 395
- DIAMOND, L K and DENTON, R L *J Lab & Clin Med.*, 1945, 30, 82
- VON DUNGERN, E and HERTZFELD, L *Ztschr f Immunit u Exp Therap* 1911, 8, 526
- ERB, I H *Univ of Toronto Med J.*, 1925, 3, 12
- FISCHER, R A, RACE, R R and associates *Nature*, 1944, 143, 106 *Ibid.*, 1945, 155, 543 *Ann Eugenics*, 1946, 13, 15
- FLOSDORF, E W, STOKES, F J and MUDD, S J *Immunol.*, 1935, 29, 389 *J Am Med Assoc.*, 1940, 115, 1095
- FREIDENREICH *Zeitsch f Immunit.*, 1931, 71, I (quoted by Weiner in *Blood groups and transfusion*, Thomas, Springfield, 1943)
- HARTMAN, F W and associates *Ann Surg.*, 1941, 114, 212
- JANEWAY, C A *New Eng J Med.*, 1941, 225, 371
- LANDSTEINER, K and LEVINE, P *J Exp Med.*, 1928, 47, 731
- LANDSTEINER, K and WEINER, A S *J Exp Med.*, 1941, 74, 309, *Proc Soc. Exp Biol Med* 1940, 43, 223, *J Exp Med* 1941, 74, 309

- LEVINE, P and associates J Am Med Assoc, 1941, 116, 825 Ann Int Med, 1941, 14, 1903 J Pediat, 1943, 23, 656 Ann N Y Acad Sc, 1946, 46, 939
- LEVINE, P and STETSON, R J Am Med Assoc, 1939, 113, 126
- PHILLIPS, R A, VAN SLYKE and associates Bull U S Army Med Service, 1943, 71, 66
- RACE, R R and associates, J Hematol, 1948, 3, 689
- STRUMIA, M M and MCGRAW, J J Ann Int Med, 1941, 15, 80
- TAYLOR, N B and WATERS, E T Can Med Assoc J, 1941, 44, 547
- WIENER, A S and associates Proc Soc Exp Biol and Med, 1946, 61, 382
- WIENER, A S and LANDSTEINER, K Proc Soc Exp Biol Med 1943, 54, 167
- WIENER, A S and PETERS, H R Ann Int Med, 1940, 13, 2306
- WITEBSKY, E and associates J Am Med Assoc, 1941, 116, 2654 Article in *Blood Substitutes and Blood Transfusions*, p 327, Thomas, Springfield, 1942 Ann N Y Acad Sc, 1946, 46, 887
- YULE, C L and associates, Blood, 1949, 4, 1232

Monographs and Reviews

- BAYLISS, W M Principles of general physiology Longmans, New York, 1918
- LATTES, L Individuality of the blood Oxford University Press, 1932
- LEVINE, P Genetic and Constitutional causes of fetal and neonatal morbidity Ann N Y Acad Sc, 1946, 46, 939
- Medical Research Council Determination of the blood groups War Memorandum No 9, London, 1943, The Rh Blood Groups, Report No 27
- MOLLISON, P L Blood Transfusion in Clinical Medicine, Blackwell, 1951
- POTTER, E L Rh—Its relation to congenital hemolytic disease and to intragroup transfusion reactions Year Book Pub Chic 1947
- WIENER, A S Blood groups and transfusion, 3rd edit Thomas, Springfield, 1943, The Rh series of genes with special reference to nomenclature, Ann N Y Acad Sc, 1946, 46, 969, An Rh glossary, Bull Dade county Med Assoc April, 1949, Rh-Hr Blood Types, Grune & Stratton, New York, 1954

CHAPTER 6

HEMOGLOBIN

- ADAIR, G S Proc Roy Soc, 1925, A 108, 627
- ANSON, M L and MIRSKY, A E J Physiol, 1925, 60, 50
- BERNHART, F W and SKEGGS, L J Biol Chem, 1943, 147, 19
- DOBRINER, K and associates J Biol Chem 1936,

- 114, XXVI Proc Soc Exp Biol Med, 1937, 36, 752, 755 Physiol Rev, 1940, 20, 416
- FISCHER, HANS Zeitsch f Physiol Chem, 1915, 95, 34 Ibid, 1915, 96, 148, Ibid, 1916, 97, 109
- GOWERS, W R Tr Chin Soc, London, 1879, xxi, 64
- HALDANE, J J Physiol, 1900-01, 25, 497
- HUFNER, G Arch Anat Physiol, 1894, p 130
- KEILIN, D Proc Roy Soc, 1926, B 100, 129
- KING, E J and associates Lancet, 1947, ii, 789, ibid, 1948, i, 282, 478, ibid 1948, ii, 563
- MEULENGRACHT, E and associates Acta med Skand, 1938, 96, 462
- NORTHROP, J H and ANSON, M L J Gen Physiol, 1929, 12, 543
- SAHLI, H Klinische Untersuchungsmethoden, Leipsic and Vienna, 5th edition, 1909, 845

Monographs and Reviews

- ANSON, M L and MIRSKY, A E Hemoglobin, the heme pigments, and cellular respiration Physiol Rev, 1930, 10, 506
- BARCROFT, J The significance of hemoglobin Physiol Rev, 1924, 4, 329 The significance of hemoglobin in submammalian forms of life Physiol Rev, 1925, 5, 596 The respiratory function of the blood Part II. Haemoglobin Cambridge University Press, 1928
- DRAKIN, D L, Metabolism of the hemin chromoproteins Physiol Rev 1951, 31, 345
- GRAY, C L, The Bile Pigments Methuen and Co, London, 1953
- MASON, V R, COURVILLE, C and ZISKIND, E Porphyrins in human disease Medicine, 1933, 12, 355
- SVEDBERG, T Molecular sedimentation in the ultracentrifuge Endeavour, 1947, 6, 89
- WATSON, C J, Lancet, Mar 10, 1951
- WATSON, C J and LARSON, E A The coproporphyrins in health and disease Physiol Rev, 1947, 27, 478

CHAPTER 7

HEMOLYSIS AND SUSPENSION STABILITY OF THE BLOOD

- BIERNAKS, E Zeitsch f Physiol Chem, 1894, 19, 179
- KURNICK, N B J Biol Chem, 1941, 140, 581
- MACGRAITH, B H and associates Lancet, 1941, 2, 530
- SIEBENS, A A and associates Blood, 1948, 3, 1367
- WALTON, A C R Quart J Med, 1935, 2, N S, 79

Monographs and Reviews

- FÄHRÆUS, R. The suspension stability of the blood Acta Med Skand, 1921, 55, 3 Physiol Rev, 1929, 9, 241

- GILLYGAN, D. R. and BLUMGART, H. L. *March Hemoglobinuria*. Medicine, 1941, 20, 341
- PONDER, E. The erythrocyte and the action of simple haemolysins Oliver, Edinburgh, 1924
- YULE, C. L. Hemoglobinuria *Physiol. Rev.*, 1942, 22, 19

CHAPTER 8

THE SPLEEN, THE LIFE OF THE RED CELL,
THE REGENERATION OF BLOOD, IRON
METABOLISM

- ASHBY, W. Determination of the length of transfused blood corpuscles in man *J. Exp. Med.* 1919, 29, 267 The life span of the red blood cells A résumé, *Blood*, 1948, 3, 486
- BALFOUR, W. M. and associates *J. Exp. Med.*, 1942, 76, 15
- BARCROFT and associates Observations upon the effect of high altitude on the physiological processes of the human body On the relation of external temperature to blood volume *Phil. Trans. Roy. Soc.*, 1922, B 211, 351 Observations on the taking up of carbon monoxide by the haemoglobin in the spleen *J. Physiol.*, 1923, 58, 138 Contribution to the physiology of the spleen *Ibid.*, 1925, 60, 443 The influence of the spleen in carbon monoxide poisoning *Ibid.*, p. 79 Rhythmical contractions of the spleen. *Ibid.*, 1932, 74, 294 Cause of rhythmical contractions of the spleen. *Ibid.*, p. 299 Undulatory changes in blood pressure. *Ibid.*, p. 311 The action of the splanchnic nerves on the spleen *Ibid.*, p. 321 Some observations on the denervated spleen *Ibid.*, 1936, 87, 189
- BARKAN, G. and WALKER, B. S. *J. Biol. Chem.*, 1940, 135, 37
- CANNON, W. B. and IZQUIERDO, J. J. *Am. J. Physiol.*, 1928, 84, 545
- COLLUMBINE, H. and SIMPSON, M. *Nature*, 1947, 159, 782
- ELVEHJEM, C. A., HART, E. B. and SHERMAN, W. C. *J. Biol. Chem.*, 1933, 103, 61
- GRANICK, S. *Science*, 1946, 103, 107, *Chem. Rev.*, 1946, 38, 379, and MICHAELIS, L. *J. Biol. Chem.*, 1943, 147, 91, *Bull. N. Y. Acad. Medicine*, 1949, 25, 403, *Physiol. Rev.* 1951, 31, 489
- GRAY, C. H. and NEUBERGER, A., *Biochem. J.*, 1949, 44, xlvii.
- GRINDLAY, J. H. and associates *Am. J. Physiol.*, 1939, 127, 119
- HAHN, P. F. *Proc. Fed. Am. Soc. Exp. Biol.*, 1948, 7, 493
- HAM, T. H. *Arch. Int. Med.*, 1939, 69, 1271
- HARGIS, E. H. and MANN, F. C. *Am. J. Physiol.*, 1925, 75, 180
- HAWKINS, W. B. and WHIPPLE, G. H. *Am. J. Physiol.*, 1938, 122, 418
- HILL, R. *Proc. Roy. Soc.*, 1930, B 107, 205

- HUGHES, J. H. and LATNER, A. L. *J. Physiol.*, 1936, 88, 388
- LANDSTEINER, K., LEVINE, P. and JAMES, M. L. *Proc. Soc. Exp. Biol. Med.*, 1928, 25, 672 See also Wiener, A. S. *J. Am. Med. Assoc.*, 1934, 102, 1779
- LEMBERG, R. and associates *Biochem. J.*, 1941, 35, 363
- McCANCE, R. A. and WIDDOWSON, E. M. *Lancet* 1937, 2, 680
- MERTENS, O. Quoted by Grindlay and associates
- SHEMIN, D. and RITTENBERG, D. *J. Biol. Chem.*, 1946, 166, 621
- WADDELL, J., STEENBOCK, H., ELVEHJEM, C. A. and HART, E. B. *J. Biol. Chem.*, 1929, 83, 251, 243
- WHIPPLE, A. O. *Ann. Surg.*, 1945, 122, 449
- WHIPPLE, G. H., HOOPER, C. W. and ROBSCHTIT, F. S. *Am. J. Physiol.*, 1920, 53, 151, 167
- WHIPPLE, G. H. and associates *Am. J. Physiol.*, 1936, 115, 651 *J. Exp. Med.*, 1937, 66, 565, *Ibid.*, 1938, 67, 89 *Ibid.*, 1939, 69, 315 *Proc. Soc. Exp. Biol. and Med.* 1937, 38, 629

Monographs and Reviews

- BARCROFT, J. Recent knowledge of the spleen *Lancet*, 1925, 1, 319
- BARCROFT, J. Features in the architecture of physiological function Cambridge University Press, 1934
- HAHN, P. F. The metabolism of iron *Medicine*, 1937, 18, 249
- PERLA, D. and MARMORSTON, J. The spleen and resistance The Williams & Wilkins Co., Baltimore, 1935
- ROBSCHTIT ROBBINS, F. S. The regeneration of hemoglobin and erythrocytes *Physiol. Rev.*, 1929, 9, 666
- ROUS, P. Destruction of the red blood corpuscles in health and disease *Physiol. Rev.*, 1923, 3, 75
- SHELDON, J. H. Haemochromatosis Oxford University Press, 1935

CHAPTER 9

THE ANEMIAS

- ANGIER, R. B. and associates *Science*, 1946, 103, 667
- VON BORNSDORFF, B. *Blood*, 1948, 3, 91
- CASTLE, W. B. and associates *Am. J. Med. Sci.*, 1929, NS 178, 764, *New Eng. J. Med.*, 1948, 239, 911, *ibid.*, 1953, 249, 603
- DAMESHEK, W. and SCHWARTZ, S. O. *Am. J. Med. Sci.*, 1938, p. 769
- FOX, H. J. and CASTLE, W. B. *Am. J. Med. Sci.*, 1942, 203, 18
- HAUSMANN, K. and MULLI, K. *Lancet*, 1952, Jan 26, p. 185
- JACOBSON, B. M. *J. Clin. Invest.*, 1935, 14, 665, 679

- JACOBSON, B M and SUBBAROW, Y J Clin Invest, 1937, 16, 573
- JUKES, T H. Folic acid and Vitamin B₁₂ in the Physiology of Vertebrates Fed Proc, 1953, 12, 633, B-Vitamins for Blood Formation, Am Lecture Series, C C Thomas, Springfield, Ill, 1952
- KARRER, P Schweiz med Woch, 1941, 71, 343
- LOUTIT, J F and MOLLISON, P L J Path Bact, 1946, 58, 711
- MINOT, G R and MURPHY, W P J Am Med Assoc, 1926, 87, 470 Ibid, 1927, 89, 759
- MITCHELL, H K, SNELL, E E and WILLIAMS, R J J Am Chem Soc, 1944, 66, 267
- MULLER, G L Am J Physiol, 1927, 82, 269
- NIEWEG, H O, cited by CASTLE, W B, New Eng J Med, 1953, 249, 603
- PFFIFFNER, J J and associates Science, 1945, 102, 228 J Biol Chem, 1945, 157, 413
- RICKES, E L and colleagues Science, 1948, 107, 396, 398 Ibid, 108, 134
- SAUBERLICH, H E and BAUMANN, G A J Biol Chem, 1948, 176, 165
- SHIVE, W and associates J Am Chem Soc, 1950, 72, 2818
- SMITH, E L Nature, 1948, 161, 638
- SPIES, T D and associates South Med J 1945, 38, 590 Blood, 1946, 1, 271 Lancet, 1948, 2, 519
- STURGIS, C C and ISAACS, R J Am Med Assoc, 1929, 93, 747
- SUÁREZ, R and associates J Lab Clin Med, 1946, 31, 2384
- THOMPSON, W P, RICHTER, M N and EDSALL, K S Am J Med Sci, 1934, NS 187, 77
- WILKINSON, J F and ISRAELS, M C G Brit Med J, 1935, 1, 139

Monographs and Reviews

- CASTLE, W B The etiology of pernicious and related macrocytic anemias Harvey Lectures, 1934-35, 30, 37
- DAVIDSON, L S P and LEITCH, I The nutritional anemias of animals and man Nutrition Abstracts and Reviews, 1934, 3, 1
- HEATH, C W and PATEK, A J The anemia of iron deficiency Medicine, 1937, 16, 267
- JOSEPHS, H W Anaemia of infancy and early childhood Medicine, 1936, 15, 307
- JUKES, T H and STOKSTAD, E L R Pteroylglutamic acid and related compounds Physiol Rev 1948, 28, 51
- PINEY, A and WYARD, S Clinical atlas of blood diseases Churchill, London, 1932
- SPIES, T D Experiences with Folic Acid The Year Book, Pub Chicago, 1947
- VAUGHN, J M The Anaemias Oxford University Press, 1934
- WELCH, A The present status of pteroylglutamic acid

and other hematopoietic agents Proc Fed Am Soc. Exp Biol, 1947, 6, 471

- WINTROBE, M M The erythrocyte in man Medicine, 1930, 9, 195
- WINTROBE, M M and BEEBE, R T Idiopathic hypochromic anemias Medicine, 1933, 12, 187

CHAPTER 10

THE WHITE BLOOD CORPUSCLES OR LEUKOCYTES THE PLATELETS

- BLACK, O A K and STANBURY, S W Lancet, 1947, I, 827
- CARREL, A J Am Med Assoc, 1924, 82, 255, J Exp Med, 1922, 36, 385
- CLARK, E R Internat Clinics, 1932, 1, Series 42, 300
- EHRICH, W E Ann N Y Acad Sc 1946, 46, 823
- HEINLE, R W and colleagues Ann Int Med, 1942, 17, 902 J Soc Exp Biol Med, 1945, 58, 5
- MALLERY, O T and McCUTCHEON, M Am J Med Sc, 1940, 200, 394
- McMASTER, P D and HUDACH, S S J Exp Med, 1935, 61, 783
- MENKEN, V Am J Path, 1943, 19, 1021 Ibid, 42, 154 Ann N Y Acad Sciences, 1946, 46, 789 Arch Path, 1946, 41, 376 Lancet, 1947, I, 660
- MILLER, F R and TURNER, D L Med Clin N Am, 1944, 28, 1376
- PAUL, J R and BRUNNELL, W W Am J Med Sc, 1932, 183, 90
- PFEIFFER, E Jahrb f Kinderheilk, 1889, 29, 257 (Quoted by Tidy, H Blood, 1948, 3, 825)
- PONDER, E and FLINN, Z M Quart J Exp Physiol, 1926, 16, 207, 277, 241
- SANDERS, A G and associates, Brit J Exp Path, 1940, 21, 254
- SPRUNT, T P and Evans, F A. Bull J Hopk Hosp 1920, 31, 410
- WHITE, A and DOUGHERTY, T F Proc Soc Exp Biol Med, 1944, 56, 26, Ann N Y Acad Sc, 1946, 46, 859

Monographs and Reviews

- BERNSTEIN, A Infectious mononucleosis Medicine, 1940, 19, 85
- DRINKER, C K and YOFFEY, J M Lymphatics, Lymph and Lymphoid tissue Harvard Univ Press, 1941
- KIRSCHBAUM, A. Recent studies on experimental leukemia Yale J Biol Med, 1944-45, 17, 163
- McCUTCHEON, M Chemotaxis in leucocytes Physiol Rev, 1946, 26, 319
- MENKIN, V The rôle of inflammation in immunity Physiol Rev, 1938, 18, 366
- TOCANTINS, L M The mammalian blood platelet in health and disease Medicine, 1938, 17, 155
- WHITE, A. Influence of Endocrine Secretions on Struc-

ture and Function of Lymphoid Tissue Harvey
Lecture, 1947-8, 43, 43

CHAPTER 11

THE ORIGIN OF THE BLOOD CELLS— HEMATOPOIESIS

- BEDSON, S P *Brit. J. Exp. Path.*, 1926, 7, 317
BOGOMOILETZ, A. A. *Ann. Rev. Soviet Med.*, 1943, 1,
101, *Brit. Med. J.*, 1943, II, 203
HOWELL, W. H. and DONAHUE, D. D. *J. Exp. Med.*,
1937, 65, 177
RICH, A. R. *Bull. Johns Hopk. Hosp.*, 1939, 65, 291
STRAUS, R. and associates *J. Immunology*, 1946, 54,
155
WATSON, C. H. *Edin. Med. J.*, 1932, 39, 229
WRIGHT, J. H. *J. Morph.*, 1910, 21, 263
YOFFEY, J. M. *J. Anat.* 1935-6, 70, 507

Monographs and Reviews

- ASCHOFF, L. *Lectures on pathology. I. Reticuloendo-
thelial system.* Hoeber, New York, 1924
DOAN, C. A. *The clinical implications of experimental
hematology.* Medicine, 1931, 10, 323
ISAACS, R. *Formation and destruction of red blood
cells.* *Physiol. Rev.*, 1937, 17, 291
JAFFÉ, R. H. *The reticulo-endothelial system in im-
munity.* *Physiol. Rev.*, 1931, 11, 277
MAXIMOW, A. A. *Relation of blood cells to connective
tissues and endothelium.* *Physiol. Rev.*, 1924, 4,
533 *The lymphocytes and plasma cells.* In Cow-
dry's *Special Cytology*, 1932, 2, 603
MAXIMOW, A. A. and BLOOM, W. *Text-book of his-
tology.* Saunders, Philadelphia, 1934
SABIN, F. R. *On the origin of the cells of the blood.*
Physiol. Rev., 1922, 2, 38 *Bone marrow.* *Physiol.
Rev.*, 1928, 8, 191
SACKS, B. *The reticulo-endothelial system.* *Physiol.
Rev.*, 1926, 6, 504
WHITNEY, C. *Hyperplasia of lymphoid tissue and
lymphocytosis.* *Medicine*, 1928, 7, 1

CHAPTER 12

THE COAGULATION CLOTTING OF BLOOD

- BARNETT, B. and MACFARLANE, R. G. *Lancet*, 1934,
2, 985
BEST, C. H., COWAN, C. and MACLEAN, D. L. *J.
Physiol.*, 1938, 82, 20
BRINKHAUSE, K. M. *Proc. Fed. Am. Soc. Exp. Biol.*,
1947, 6, 389
BRODIE, T. G. and RUSSELL, A. E. *J. Physiol.*, 1897,
21, 403
CANNON, W. B. and MENDENHALL, W. L. *Am. J.
Physiol.*, 1914, 34, 225
CHAPGAFF, E. and OLSON, K. B. *J. Biol. Chem.*, 1937,
122, 153
CHARLES, A. F. and SCOTT, D. A. *J. Biol. Chem.* 1933,
102, 425, 437 *Biochem. J.*, 1936, 30, 1927

- DRINKER, C. K. and DRINKER, K. R. *Am. J. Physiol.*,
1910, 41, 5
FANTL, P. and NANCE, M. *Nature*, 1946, 158, 708
FERGUSON, J. H. *Am. J. Physiol.*, 1937, 119, 755
Proc. Soc. Exp. Biol. Med., 1939, 42, 33
HEDENIUS, P. *Lancet*, 1937, 2, 1186
HOWELL, W. H. and HOLT, E. *Am. J. Physiol.*, 1918,
47, 328
IVY, A. C., SHAPIRO, P. F. and MELNICK, P. *Surg.
Gynec. and Obst.*, 1935, 60, 781
JAQUES, L. and associates, *Can. Med. Assoc. J.*, 146,
25, 26
JORGES, E. *Naturwissenschaften*, 1935, 23, 196
DE JUNG, B. *La Coagulation.* Herman, Paris, 1936
LINK, K. P. and associates *J. Biol. Chem.*, 1941, 138,
21 *Ibid.*, 1943, 147, 463 *Harvey Lectures* 1943-
44, 39, 162 *Proc. Fed. Am. Soc. Exp. Biol.*, 1944,
4, 176
MASON, E. C. *Surg. Gynec. and Obst.*, 1924, 39, p. 421
MCLEAN, J. *Am. J. Physiol.*, 1916, 41, 250
MILLS, C. A. *Chinese J. Physiol.*, 1927, 1, 245
MOMMAERTZ, W. F. H. M. *J. Gen. Physiol.*, 1946, 29,
103, 113
MOOLTON, S. E. and associates, *Arch. Int. Med.*, 1949,
84, 667
MURRAY, D. W. G., JAQUES, L. B., PERRETT, T. S.
and BEST, C. H. *Surgery*, 1937, 2, 163
OCHSNER, A. *Surgery*, 1945, 17, 240
OWREN, P. A. *Thesis*, Oslo, 1947, *Parahemophilia*,
Lancet, 1947, I, 446
PICKERING, J. W. *The influence of Witte's "peptone"
and of digestion of blood platelets and plasma.*
Proc. Roy. Soc., 1929, B 104, 512
PINNINGER, J. L. and PRUNTY, F. T. G. *Brit. J. Exp.
Pathol.*, 1946, 27, 200
QUICK, A. J. *On the action of heparin and its relation
to thromboplastin.* *Am. J. Physiol.*, 1936, 115,
317 *Am. J. Med. Sci.*, 1941, 201, 469 *On the
constitution of prothrombin.* *Am. J. Physiol.*,
1943, 140, 212, *ibid.*, 1943, 146, 212 *On the quan-
titative estimation of prothrombin.* *Am. J. Clin.
Path.*, 1945, 15, 560 *The components of prothrom-
bin.* *Proc. Soc. Exp. Biol. Med.*, 1946, 62, 249
*Congenital hypoprothrombinemia and pseudo-
prothrombinemia.* *Lancet*, 1947, II, 379 *Coagula-
tion factors.* *Ibid.*, 1947, II, 772 *Components of
the prothrombin complex.* *Am. J. Physiol.*, 1947,
151, 63
RODERICK, L. M. *J. Vet. Med. Assoc.*, 1929, 74, 314
ROUS, P. and associates *J. Exp. Med.*, 1930, 51, 807
SANDISON, J. C. *Anat. Rec.*, 1931, 60, 355
SCHOFIELD, F. W. *Can. Vet. Rec.*, 1922, 3, 74 *Am.
Vet. Med. Assoc.*, 1924, 64, 553
SEEGER, W. H. and associates *J. Biol. Chem.*, 1938,
123, 751 *Science*, 1947, 106, July 11 *Am. J.
Physiol.*, 1947, 150, 58 *Ibid.*, 1948, 152, 567 *J.
Biol. Chem.*, 1948, 174, 565, *Am. J. Clin. Path.*,
1949, 19, 41, *J. Biol. Chem.*, 147, 169, 231

- TOCANTINS, L M Platelets and the structure and physical properties of blood clots *Am J Physiol*, 1936, 114, 709 The bleeding time *Am J Clin Path*, 1936, 6, 160 Demonstration of antithromboplastic activity in normal and hemophilic plasmas *Am J Physiol*, 1943, 139, 265 Relation of contacting surfaces and anticephalin activity with reference to the fluidity and coagulability of the blood *Blood*, 1946, 1, 156
- TROLAND, C E and LEE, F C *Bull J Hopk. Hosp*, 1937, 62, 85
- WARNER, E D and associates *Arch Pathol*, 1934, 18, 587 *Am J Physiol*, 1936, 114, 667 *J Exp Med*, 1937, 66, 801
- WATERS, E T, MARKOWITZ, J and JAUQUES, L B *Science*, 1938, 87, 582
- WIDSTRÖM, G and WILANDER, O *Acta Med Scandinav*, 1936, 88, 434.
- WILANDER, O *Skand Arch Physiol*, 1938, 81, Supp 15, p 89
- WRIGHT, H. P *J Path Bact*, 1942, 54, 461
- LEVY, R. L., ROWNTREE, L G and MARRIOTT, W M *Arch Int Med*, 1915, 16, 309
- VAN SLYKE, D D The normal and abnormal variations in the acid-base balance of the blood *J Biol Chem* 1921, 48, 153 A method for the determination of carbon dioxide and carbonates in solution *Ibid*, 1917, 30, 347 VAN SLYKE, D D and CULLEN, G E The bicarbonate concentration of blood plasma, its significance and its determination as a measure of acidosis *J Biol. Chem*, 1917, 30, 289

Monographs and Reviews

- HENDERSON, L J *Blood, a study in general physiology* Yale University Press, 1928
- OGSTON, A. G The definition and meaning of pH *Physiol Rev*, 1947, 27, 228
- ROUGHTON, F J W Recent work on carbon dioxide transport by the blood *Physiol Rev*, 1935, 15, 241

SECTION II

CHAPTER 14

THE DYNAMICS OF THE CIRCULATION

- EVANS, C L *J Physiol*, 1918, 52, 6 EVANS, C L and MATSUOKA, Y *J Physiol*, 1915, 49, 378.

CHAPTER 15

THE ARTERIAL BLOOD PRESSURE

- BAZETT, H. C and associates *Am J Physiol*, 1935, 113, 312
- GREEN, H. E *Am J Physiol*, 1944, 141, 518
- HALLOCK, P and BENSON, J C *J Clin Invest*, 1937, 16, 595
- WOODBURY, R. A and HAMILTON, W F *Am J Physiol*, 1937, 119, 663
- ASCHOFF, L *Lectures on pathology XI. Thrombosis* Hoeber, New York, 1924
- EAGLE, H Recent advances in blood coagulation problem *Medicine*, 1937, 16, 95
- FERGUSON, J H. The blood calcium and the calcium factor in blood coagulation *Physiol Rev*, 1936, 16, 640 Mechanism of blood coagulation *Am J Med*, 1947, 3, 67 A review of some basic facts of blood coagulation *Ann N Y Acad Sc.*, 1947,
- HOWELL, W H. The coagulation of blood *Harvey Lectures*, 1916-17, 12, 272 Theories of blood coagulation *Physiol Rev*, 1935, 15, 435 A textbook of physiology Saunders, Philadelphia, 1933 Hemophilia W M Carpenter Lecture, *Bull N York Acad Med*, 1939, 15, 3 Thromboplastic substances *Bull J Hopk. Hosp*, 1945, 76, 295
- JORPES, J E *Heparin* Oxf Univ Press Second Edition, 1949
- PICKERING, J W The blood plasma in health and disease Heinemann, London, 1928
- QUICK, A. J The hemorrhagic diseases, Thomas, Springfield, 1942 Anticoagulants *Physiol Rev*, 1944, 24, 297, The Physiology and Pathology of Hemostasis Lea and Febiger, Phila., 1951
- SILBERBERG, M The causes and mechanism of thrombosis *Physiol Rev*, 1938, 18, 197
- WELCH, W H. Thrombosis, In Allbutt and Rolleston's *System of Medicine*, 1909, 6, 691

CHAPTER 16

THE BLOOD PRESSURE IN THE HUMAN SUBJECT THE VENOUS BLOOD FLOW

- ABRAMSON, D I *Proc. Soc. Exp Biol Med*, 1940, 45, 127
- ANDERSON, H. *Arch Int Med*, 1926, 37, 297
- ANSELMINO, K J and HOFFMANN, F *Klin Wchnschr*, 1931, 10, 1438
- APPERLY, F L and CARY, M K *Proc Soc Exp Biol. Med*, 1941, 48, 492
- ARMSTRONG, H G *Am J Med Sci*, 1938, 195, 211
- ARNOTT, W M and KELLEP, R J *J Path Bact*, 1936, 42, 141
- BAYLISS, W M and STARLING, E H *J Physiol*, 1894, 16, 159
- BAZETT, H C *Am J Physiol*, 1924, 70, 550
- BOHNING, A., JOCHIM, K and KATZ, L N *Am J Physiol*, 1933, 106, 183

CHAPTER 13

THE MECHANISMS REGULATING THE REACTION OF THE BODY FLUIDS

- CULLEN, G E *J Biol Chem*, 1922, 50, xvii
- DALE, H. H. and EVANS, C L *J Physiol*, 1920, 54, 157

- BRAUN-MENENDEZ, E Hypertension after bilateral nephrectomy *Nature*, 1947, 160, 905
- BRAUN-MENENDEZ, E and associates *J Physiol*, 1940, 98, 283 *Rev Soc. Argent Biol*, 1940, 16, 643
- BRIGHT, R. Guy's Hosp Rep, 1836, 1, 338 *Reports Med Cases, Lond*, 1827
- BYROM, F B and WILSON, C *Quart J Med*, 1941, 10, 65
- CASTELMAN, B and SMITHWICK, R. H. *J Am Med Ass*, 1943, 121, 1256
- COLES, B C, RICHARDSON, H F AND HALL, G E *Can Med Assoc J*, 1937, 37, 24
- CORCORAN, A C and PAGE, I H. *Am J Physiol*, 1938, 123, 43 *Ibid*, 1942, 135, 361
- COTTON, T F, RAPPORT, D L and LEWIS, T *Heart*, 1917, 6, 269
- CROVATTO, H. and associates *Science*, 1942, 95, 101
- Ibid*, 1943, 96, 519 *Proc Soc. Exp Biol and Med*, 1943, 52, 61
- CUSHING, H *Am J Med Sci*, 1902, 124, 375
- DAWSON OF PENN (LORD) Discussion on hyperpiesia *Brit Med J*, 1925, 2, 1161
- DEXTER, L and HAYNES, W *Proc Soc Exp Biol and Med*, 1944, 55, 288
- DILL, L V and ERICKSON, C C *Proc Soc. Exp Biol and Med*, 1938, 39, 362
- DIXON, W E and HELLER, H *Arch f Exp Path u Pharmacol*, 1932, 166, 265
- DOCK, W *Am J Physiol*, 1940, 130, 1, *J Clin Invest*, 1940, 19, 769
- DRURY, D R. *J Exp Med*, 1938, 68, 693
- DUFF, G L and associates *Proc Soc Exper Biol and Med*, 1939, 41, 295
- EDMAN, P *Arkiv f Kemi Mineralogi och Geol*, Stockholm, 1945, 12A, 1
- ELLIS, L B, and WEISS, S *J Clin Invest*, 1930, 8, 47
- ERLANGER, J AND HOOKER, D R *Johns Hopkins Hosp Rep*, 1904, 12, 145
- GEER, W A and DRAGSTEDT, L R *Ann. Surg*, 1938, 108, 263
- GOFMAN, J W, in *Factors Regulating Blood Pressure*, 5th Conference, J Macy Jr Found, 1951, pag 46
- Hypertension, a Symposium, Univer Minnesota Press, 1951
- GOLDBLATT, H *Factors Regulating Blood Pressure*, J Macy Jr Found, 5th Conference, 1951
- GOLDBLATT, H, LYNCH, J, HANZAL, R F, and SUMMERVILLE, W W *J Exp Med*, 1934, 59, 347
- GREENWOOD, W F, NASSIM, R. and TAYLOR, N B *Can Med Assoc J*, 1939, 41, 443
- GOLDRING, W and associates *J Clin Invest*, 1941, 20, 637
- GRIMSON, K S *Ann Surg*, 1941, 114, 753
- GROLLMAN, A, HARRISON, T R AND WILLIAMS, J R *Amer Assoc. Adv Sci.*, Pub 13, 1940, p 239
- J Am Med Assoc*, 1940, 115, 1169
- HELMER, O M and PAGE, I. H. *J Biol Chem*, 1939, 127, 757
- HILL, L AND BARNARD, H *J Physiol*, 1897, 21, 323
- HILL, L and ROWLANDS, R A *Heart*, 1912, 3, 219
- HINES, E A, JR. *Am Heart J*, 1940, 19, 408
- HOOKE, D R. *Am J Physiol*, 1911, 28, 235
- KATZ, L N and colleagues *Am. Heart J*, 1939, 17, 334
- KOHLSTAEDT, K G and PAGE, I H. *J Exp Med*, 1940, 72, 201
- HOUSAY, B A, FASCILOLO, J C and TAQUINI, A C *J Physiol*, 1938, 94, 281
- JOSUÉ, O *Compt rend soc biol*, 1903, 55, 1374
- KATZ and associates, see Bohning, Jochim and Katz
- LEVY, S E and his associates *Am J Physiol*, 1938, 122, 38
- MACWILLIAM, J A *Quart J Exp Physiol*, 1923, 13, 178 *Brit Med J*, 1923, 2, 1196
- MILLER, J H and BRUGER, M *Am Heart J*, 1939, 18, 329
- MORITZ, A R and OLDT, M R. *Am J Path*, 1937, 13, 679
- MYER, E and colleagues *Am J Physiol*, 1947, 152, 397
- OGDEN, G and associates Rise in blood pressure during ischemia of the gravid uterus *Proc Soc. Exp Biol and Med*, 1940, 43, 49
- OLDT, see MORITZ and OLDT
- PAGE, E W, OGDEN, E and associates *Proc Soc Exp Biol and Med*, 1940, 43, 49 *Am J Physiol*, 1944, 141, 707 *Surg Gynec and Obstet*, 1943, 78, 493
- PAGE, I H. and associates *Am J Physiol*, 1937, 120, 238 *J Am Med Assoc*, 1939, 113, 2046 *J Exp Med*, 1941, 73, 7 *Ann Int Med*, 1941, 15, 347
- PEET, M M and ISBERG, E M *J Amer Med Assoc.*, 1946, 130, 467
- PICKERING, G W *Clin Sci* 1936, 2, 209, 363
- ROBINSON, S C N *Eng J Med*, 1940, 223, 407
- ROBINSON, S C and BRUCER, M *Arch Int Med*, 1939, 64, 409
- SCHROEDER, H A *Science*, 1942, 95, 306
- SCHROEDER, H A. and ADAMS, M H *J Exp Med*, 1941, 73, 531
- SHORR, E in *Hypertension*, a Symposium, BELL, E T, Ed Univer Minnesota press, 1951, page 79, ZWEIFACH, B W and SHORR, E *Factors Regulating Blood Pressure J Macy Jr Found*, 5th Conference, 1951
- SKELTON, R *J Physiol*, 1921, 55, 319
- SOLANDT, D Y, NASSIM, R. and COWAN, C R *Lancet*, 1940, 1, 873
- STARLING, E H. Discussion on hyperpiesia *Brit Med J*, 1925, 2, 1163
- SYMONDS, B *J Am Med Assoc*, 1923, 80, 232
- TAYLOR, N B and MACLACHLAN, I *Am J Physiol*, 1940, 120, 413

- THEOBALD, G W *Clinical Sci*, 1934, 1, 225
 THOMAS, C B J *Hopk Hosp Bull*, 1934, 74, 335
 TIGERSTEDT, R and BERGMANN, P G *Skand Arch Physiol*, 1898, 8, 223
 TRUETA, J and associates *Studies of the Renal circulation* Thomas, Springfield, 1947
 VERNEY, E B and VOGT, M *Quart J Exp Physiol*, 1938, 28, 253
 WAKERLIN, G E and associates *Am J Physiol*, 1950, 163, 701
 WEISS, S *New Eng J Med*, 1930, 202, 891
 WILSON, C and BYROM, F B, *Lancet*, 1939, 1, 136

Monographs and Reviews

- ALLBUTT, T C *Diseases of the Arteries including Angina Pectoris* Macmillan, London, 1915
 BELL, E T, Ed *Hypertension, a Symposium*, Univ Minnesota Press, Minneapolis, 1951
 EYSTER, J A E *Venous pressure and its clinical applications* *Physiol Rev*, 1926, 6, 281 *The clinical aspects of venous pressure* Macmillan, New York, 1929
 FRIEDLANDER, A *Hypotension* Medicine, 1927, 6, 143
 HEYMANS, C *The Regulation of Blood Pressure and Heart Rate*, Amer Lecture Series, No 43, C C Thomas, Springfield, 1950
 GOLDBLATT, H *Harvey Lectures*, 1937-8, p 237
 GOLDRING, W and CHASSIS, H. *Hypertension and hypertensive disease* Commonwealth Fund, 1944
 LEWIS, H A and GOLDBLATT, H *Studies on experimental hypertension* *Bull N Y Acad Med*, 1942, 18, 459
 MACWILLIAM, J A *Blood pressures in man under normal and pathological conditions* *Physiol Rev*, 1925, 5, 303
 PAGE, I H *Special aspects of the problem of renal origin of hypertension* *Bull N Y Acad Med*, 1943, 19, 461
 WEISS, S *The etiology of arterial hypertension* *Ann Int Med*, 1934, 8, 296
- The thermo-stromuhr method of measuring blood flow *Physics*, 1931, 1, 407
 HEWLETT, A W and VAN ZWALUWENBURG, J G *Heart*, 1909, 1, 87
 REIN, H Z *Biol*, 1928, 87, 394
 ROBB, G P and WEISS, S *Am Heart J*, 1933, 8, 650
 SMITH, L A and ALLEN, E V *Staff Meet Mayo Clin*, 1941, 16, 53
 STEWART, G N *Studies on the circulation in man* *Harvey Lectures*, 1912-13, 8, 86 See also, BROWN, G E *Calorimetric studies of the extremities* *J Clin Invest*, 1926, 3, 369, KEGGERREIS, R *Calorimetric studies of the extremities (Stewart's method)* *J Clin Invest*, 1926, 3, 357 SHEARD, C *Calorimetric studies of the extremities (Stewart's method)* *J Clin Invest*, 1926, 3, 327 TAYLOR, N B *Blood-flow in man as estimated by the calorimetric method of Stewart* *J Lab Clin Med*, 1922, 7, 430 *Estimations of the blood-flow through the hands in clinical cases* *J Lab Clin Med*, 1922, 8, 153
 WOOD, E H and associates *Effects of acceleration in relation to aviation* *Proc Fed Am Soc Exp Biol*, 1946, 5, 327

Monograph

- ARMSTRONG, H G *Principles and practice of aviation medicine* The Williams & Wilkins Co, Baltimore, 3rd. ed, 1952

CHAPTER 18

THE ARTERIAL PULSE

- BRAMWELL, J C, HILL, A V and MCSWINEY, B A *Heart*, 1923, 10, 233 See also, BRAMWELL, J C and HILL, A V *Proc Roy Soc*, 1922, B 93, 298 HICKSON, S K and MCSWINEY, B A *J Physiol*, 1924, 59, 217

Monographs and Reviews

- HILL, L *The mechanism of the circulation of the blood* In Schäfer's *Text-book of Physiology*, 1900, 2, 1
 WIGGERS, C J *Modern aspects of the circulation in health and disease* Lea, Philadelphia

CHAPTER 19

THE PHYSIOLOGY OF CARDIAC MUSCLE PERFUSION OF THE ISOLATED HEART

- ANDRUS, E C and CARTER, E P *Heart*, 1924, 11, 97
 CLARK, A. J *J Physiol*, 1913, 47, 66, *J Physiol*, 1931, 75, 25P
 CRUICKSHANK, E W H *Cardiac metabolism* *Physiol Rev*, 1936, 16, 597
 CRUICKSHANK, E W H and KOSTERLITZ, H W *J Physiol*, 1941, 99, 208
 DRURY, A N *Quart J Exp Physiol*, 1940, 30, 59

CHAPTER 17

THE VELOCITY OF THE BLOOD

- BLUMGART, H L and YENS, O C *J Clin Invest.*, 1927, 4, 1
 BRODIE, T G and RUSSELL, A E *J Physiol*, 1905, 32, xlvii
 BURTON, A C, *American Jour Physiol*, 1939, 127, 437
 BURTON-OPITZ, R *Am J Physiol*, 1903, 9, 198
 GREEN, H D and associates *Am J Physiol*, 1940, 129, 366 (Proc)
 HERRICK, J F, ESSEX, H E and BALDES, E J *The effect of lumbar sympathectomy on the blood flow in the femoral artery of the dog* *Am J Physiol*, 1932, 101, 213 HERRICK, J F and BALDES, E J

- EVANS, C L AND STARLING, E H. *J Physiol*, 1914, 49, 67
- EVANS, C L and associates *J Physiol*, 1933, 80, 21
Quart J Exp Physiol., 1935, 24, 347
- FIDDES, J. *Quart. J Exp Physiol.*, 1929, 19, 262, 265
- FLETCHER, J P and WATERS, E T. *J Physiol*, 1938, 84, 337
- KATZ, L N and LONG, C N H. *Proc. Roy Soc.*, 1926, B 99, 8

KEITH, N M. *Ann Int. Med.*, 1942, 16, 879
Staff Meet Mayo Clinic, 1946, Oct. 2, 392

MANN, F C and associates *Arch Surg*, 1933, 26, 219

McGINTY, D A and MILLER, A T, JR. *Am J Physiol*, 1932, 101, 76 (*Proc.*) *Ibid*, 1933, 103, 712
See also, HIMWICH, H. E., KOSKOFF, Y D and NAHUM, L. H. Changes in lactic acid and glucose in the blood on passage through organs
Proc. Soc. Exp Biol Med, 1928, 26, 347
McGINTY, D A. Studies on the coronary circulation
I Am J Physiol, 1934, 98, 244.

PLINGER, S J. *Physiol*, 1883, 4, 29

TAYLOR, N B. *J Physiol*, 1931, 71, 156

VISSCHER, M B. *Proc. Soc. Exp Biol and Med*, 1938, 38, 323

WÉGRIA, R, MOE, G K and WIGGERS, C J. *Am J Physiol.*, 1941, 133, 657

WIGGERS, see WÉGRIA, MOE and WIGGERS

WOODWORTH, R S. *Am J Physiol*, 1902, 8, 213

Reviews

- CLARKET, A J, EGGLETON, M G et al. The metabolism of the frog's heart. Oliver and Boyd, Edin, 1938
- CRICKSHANK, E W H and McCLEURE, G S. *J Physiol*, 1936, 86, 1
- MEEK, W J. The question of cardiac tonus. *Physiol. Rev.*, 1927, 7, 259

CHAPTER 20

THE ORIGIN AND MODE OF TRANSMISSION OF THE HEART BEAT

- ABRAMSON, D I and associates. *J Anat.*, 1936, 70, 250
Am J Physiol, 1937, 120, 635
- BARKER, P S, MACLEOD, A G and ALEXANDER, J. *Am Heart J*, 1930, 5, 729
- COHN, A E, KESSEL, L and MASON, H. H. *Heart*, 1912, 3, 311
- ERLANGER, J and BLACKMAN, J R. *Am J Physiol*, 1907, 19, 125
Chronic Heart, 1910, 1, 177
- GASKELL, W H. In Schäfer's *Textbook of Physiology* Pentland, Edinburgh, 1900, 2, 169
- KEITH, A and FLACK, M. *J Anat.*, 1907, 41, 172
- KEITH, A. and MACKENZIE, I. *Lancet*, 1910, 1, 101
- KENT, A F S. *J Physiol.*, 1893, 14, 233
- LEWIS, T and ROTHSCHILD, M A. *Phil. Trans.*, London, Series B, 1915, 206, 181
- ROBB, J S and ROBB, R. C. *Am J Physiol.*, 1936, 116, 43

Monographs and Reviews

- ERLANGER, J. The localization of impulse initiation and conduction in the heart. *Harvey Lectures*, 1912-13, 8, 44
- FLACK, M. The heart. In *Further Advances in Physiology* Arnold, London, 1909, p 34
- LEWIS, T. *Lectures on the heart*. Hoeber, New York, 1915
- LEWIS, T. The mechanism and graphic registration of the heart beat. Shaw, London, 1925

CHAPTER 21

THE ARCHITECTURE OF THE HEART, MOVEMENTS OF THE HEART, THE CARDIAC CYCLE

- GOULEY, B A. *Am Heart J*, 1941, 22, 208
- HENDERSON, Y. *Am J Physiol*, 1906, 16, 325
- WIGGERS, C J. *Am J Physiol*, 1916, 40, 218

Monographs and Reviews

- PRINZMETAL, M and associates. *Auricular Arrhythmias* C C Thomas, Springfield, Ill, 1952
- WIGGERS, C J. The pressure pulses in the cardiovascular system. Longmans, New York, 1928
Modern aspects of the circulation in health and disease. Lea, Philadelphia, 1923

CHAPTER 22

MOVEMENTS OF THE HEART VALVES, THE HEART SOUNDS

- DEAN, A L, JR. *Am J Physiol*, 1916, 40, 206
- DOCK, W. *Arch Int Med*, 1933, 51, 737
- HENDERSON, Y and JOHNSON, F E. *Heart*, 1912, 4, 69
- KOUNTZ, W B and associates. *Am Heart J*, 1940, 20, 667
- LEWIS, T. Illustrations of heart sound records. *Quart J Med*, 1913, 6, 441
- STEWART H. A. *Arch. Int Med*, 1908, 1, 100
- STRAUB, D. *Dtsch Arch. Klin Med*, 1914, 116, 409
- WIGGERS, C J. *Arch Int Med*, 1919, 24, 471
J Am Med Assoc, 1931, 97, 1359
- WIGGERS, C J and DEAN, A L, JR. *Am J Med Sci*, 1917, 153, 666
Am J Physiol, 1917, 42, 476

Monographs and Reviews

- LEWIS, T. Diseases of the heart, described for practitioners and students. Macmillan, London, 1933
- ORIAS, O and BRAUN-MENENDEZ, E. The heart sounds in normal and pathological conditions. Oxford Univ Press, 1939
- WHITE, P D. *Heart disease*. Macmillan, New York, 1931

CHAPTER 23

ELECTROCARDIOGRAPHY THE VENOUS PULSE

- BARKER, P S, MACLEOD, A G and ALEXANDER, J. *Am Heart J*, 1930, 5, 720

- BATTRO, A AND BIDOGGIA, H *Am Heart J*, 1947, **33**, 604
- CRAIB, W H *Heart*, 1927, **14**, 71
- GOTCH, F *Heart*, 1910, **1**, 235
- HOFF, H S and NAHUM, L H *Am J Physiol*, 1941, **131**, 700
- LEWIS, T The spread of the excitation process in the vertebrate heart *Phil Trans Roy Soc*, 1916, **B 207**, 221 The interpretation of the initial phases of the electrocardiogram with special reference to the theory of "limited potential differences" *Arch Int Med*, 1922, **30**, 269
- WALLER, A D *J Physiol*, 1879-1880, **2**, 384
- WILSON, F N and FINCH, R *Heart*, 1923, **10**, 275
- WOLFERTH, C C and WOOD, F C *Am J Med Sci*, 1932, **183**, 30
- MINES, G R *J Physiol*, 1913, **46**, 349 *Trans Roy Soc Can*, 1914, Sec IV, **8**, 43
- NAHUM, L H and HOFF, E H *J Pharmac Exp Therap*, 1939, **65**, 322
- PRINZMETAL, M and associates *The Auricular Arrhythmias* C C Thomas, Springfield, Ill., 1952, Circul, 1950, **1**, 341, *J Am Med Assoc*, 1951, **46**, 275
- ROBERTS, G H and associates *Am Heart J*, 1932, **7**, 505 *J Clin Invest*, 1935, **14**, 867
- WIGGERS, C J *Am J Physiol*, 1930, **92**, 223 *Am Heart J*, 1940, **20** 399, 413
- WILSON, F N, MACLEOD, A G and BARKER, P S *Am Heart J*, 1931, **6**, 637
- WOOD, F C and WOLFERTH, C C *Am Heart J*, 1943, **25**, 454

Monographs and Reviews

- CRAIB, W H *The electrocardiogram* Med Res Coun, London, Spec Rep No 147, 1930
- GOLDBERGER, E *Unipolar Electrocardiography*, Lea and Febiger, Phila, 1949
- HECHT, H H *Basic Principles of Clinical Electrocardiography*, Am Lecture Series C C Thomas, Springfield, Ill, 1950
- KATZ, L N *The genesis of the electrocardiogram* *Physiol Rev*, 1947, **27**, 398 *Electrocardiography* Lea and Febiger, Phila, 1941

CHAPTER 24

DISORDERS OF THE HEART BEAT AND THEIR INVESTIGATION BY GRAPHIC METHODS

- BARKER, P S, MACLEOD, A G and ALEXANDER, J *Am Heart J*, 1930, **5**, 720
- BRAMS, W A and KATZ, L N *Am Heart J*, 1931, **7**, 249
- BUTTERWORTH, J S and POINDEXTER, C A *Arch Int Med*, 1942, **69**, 437
- CATTELL, MCK and GOLD, H *Arch Int Med*, 1940, **65**, 263 *J Pharm Exp Therap*, 1938, **62**, 116
- GARREY, E W *Am J Physiol*, 1914, **33**, 397
- HEYMANS, C and associates *Compt rend Soc Biol*, 1932, **110**, 572
- HOOKE, D R *Am J Physiol*, 1929, **91**, 305 *Ibid*, 1930, **92**, 639 *Ibid*, 1932, **99**, 279 *Ibid*, 1933, **103**, 444
- KATZ, L N and associates *Am Heart J*, 1938, **16**, 149
- KEHR, N D and HOOKE, D R *Am J Physiol*, 1935, **112**, 301
- LEVY, A G *Heart*, 1913, **4**, 319 *Ibid*, 1914, **5**, 299 *Ibid*, 1919, **7**, 105 *Brit Med J*, 1914, **2**, 502
- MACWILLIAM, J A *J Physiol*, 1887, **8**, 296 *Proc Roy Soc*, 1918, **B 90**, 302
- MAYER, A G *Papers from Tortugas Laboratory, Carnegie Inst*, Washington, 1908, **1**, 115

Monographs and Reviews

- GARREY, W E *Auricular fibrillation* *Physiol Rev*, 1924, **4**, 215
- LEWIS, T *Clinical disorders of the heart beat* Hoeber, New York, 1916
- LEWIS, T *The mechanism and graphic registration of the heart beat* Shaw, London, 1925
- LEWIS, T *Clinical science, illustrated by personal experiences* Shaw, London, 1934
- MACKENZIE, J *Diseases of the heart* Frowde, London, 1913
- PARDEE, H E B *Clinical aspects of the electrocardiogram* Hoeber, New York, 4th Ed, 1941
- ROTH, I R *Cardiac arrhythmias, clinical features and mechanism of the irregular heart* Hoeber, New York, 1928
- WOLLENBERGER, A *Cardiac energy metabolism* *Pharm Rev*, 1949, **1**, 311
- WILLIUS, F A *Clinical electrocardiograms Their interpretation and significance* Saunders, Philadelphia and London, 1929

CHAPTER 25

THE REGULATION OF THE HEART'S ACTION, HEART FAILURE

- ALAM, M and SMIRK, F H *J Physiol*, 1938, **92**, 16u
- ANREP, G V and SEGALL, H N *Heart*, 1925, **13**, 61
- ASHMAN, R and GARREY, W E *Am J Physiol*, 1931, **98**, 109
- BAIN, W A *Quart J Exp Physiol*, 1932, **22**, 269
- BAINBRIDGE, F A *J Physiol*, 1914, **48**, 332 *Ibid*, 1915, **50**, 65
- BEATTIE, J, BROW, G R and LONG, C N H *Proc Roy Soc*, 1930, **B 106**, 253
- BOND, D D *Am J Physiol*, 1943, **138**, 468
- BOWEN, W P *Contributions to Medical Research Dedicated to Victor Vaughan*, 1903, p 462 *Am J Physiol*, 1904, **11**, 59
- BRINKMAN, R and VAN DAM, E *Pflüger's Arch*, 1922, **196**, 66

- BRONK and associates *Am J Physiol*, 1936, 117, 237
- BUCHANAN, F *Trans Oxford Univ Scientific Club*, 1909, 34, 351
- CANNON, W B, BRITTON, S W and LEWIS, J T *Am J Physiol*, 1926, 77, 326
- CHAPMAN, E M and assoc *Journ Am Med Assoc.*, 1948, 137, 579
- ENGELMANN, TH W *Arch f Physiologie*, 1900, p 313 and 1902, suppl, p 1
- v EULER, U S *J Physiol*, 1946, 105, 39
- EYSTER, J A E *Trans Assoc Am Physic*, 1927, 42, 15 *J Am Med Assoc*, 1928, 91, 1881 *Am J Physiol*, 1930, 93, 647 (Proc.) EYSTER, J A E, MEEK, W J and HODGES, F J *Arch Int Med*, 1927, 39, 536
- GASSER, H S and MEEK, W J *Am J Physiol*, 1914, 34, 48
- GREEN, H D and HOFF, E C *Am J Physiol*, 1936, 117, 411 *Ibid*, 1937, 118, 641
- HARRISON, T R Arterial and venous pressure factors in congestive heart failure *Physiol Rev*, 1938, 18, 86
- HENDERSON, Y Acapnia and shock I Carbon-dioxide as a factor in the regulation of the heart rate *Am J Physiol*, 1908, 21, 126 III Shock after laparotomy, its prevention, production and relief *Ibid*, 1909, 24, 66 Acapnia as a factor in post operative shock, atelectasis and pneumonia *J Am Med Assoc*, 1930, 95, 572 The volume of the circulation and its regulation by the venopressor mechanism *Ibid*, 1931, 97, 1265
- HENDERSON, Y, HAGGARD, H W and CORBURN, H C The therapeutic use of carbon dioxide after anesthesia and operation *J Am Med Assoc*, 1920, 74, 783
- HOOKE, D R *Am J Physiol*, 1907, 19, 417
- HOWELL, W H and DUKE, W W *Am J Physiol*, 1908, 21, 51
- KATZ, L N Observations on cardiac failure and the mode of its production *Am Assoc Adv Science*, 1940, No 13, p 184, *Am J Physiol*, 1928, 87, 348
- KATZ, L and associates *Arch Int Med*, 1934, 63, 710
- KNOWLTON, F P and Starling, E H *J Physiol*, 1912, 44, 206
- KORTH, C *Ann Int Med*, 1937, 11, 492
- LANDIS, E J *Clin Invest*, 1925, 46, 237
- LEWIS, T and COHN, A E *J Exp Med*, 1913, 18, 739
- LINDHARD, J *Pflüger's Arch*, 1915, 161, 233
- LOEWI, O Übertragbarkeit der Herznervenwirkung I Mitteilung *Pflüger's Arch*, 1921, 189, 239
- LOEWI, O and NAVRATIL, E Über humorale Übertragbarkeit der Herznervenwirkung X Mitteilung Über das Schicksal des Vagusstoffes *Pflüger's Arch*, 1926, 214, 687 XI Mitteilung Über den Mechanismus der Vaguswirkung von Physostigmin und Ergotamin *Ibid*, 689
- LUCKE, H *Deutsch Arch f Klin Med*, 1937, 180, 40
- MCDONALL, R. J S A vago-pressor reflex *J Physiol*, 1924, 59, 41 A right auricular pressor reflex *Ibid*, 1934, 81, 5 P Acetylcholine stimulation of the heart *J Physiol*, 1946, 104, 392
- MERRILL, A J *J Clin Invest*, 1916, 25, 369, *ibid*, 1949, 26, 389, *Am J Med*, 1949, 6, 357
- MILLER, F R and BOWMAN, J T *Am J Physiol*, 1916, 39, 149
- MITCHELL, Quoted by Allbutt in Allbutt and Rolleston's *System of Medicine* Macmillan, 1909, 6, 199
- MOORE, R M and CANNON, W B *Am J Physiol*, 1930, 94, 201
- NAIUM, L H and HOFF, E C *Am J Physiol*, 1935, 113, 101
- NONDEZ, J F *Am J Anat*, 1939, 65, 361 *Am Heart J*, 1943, 26, 577
- OGDEN, E and SHOCK, N W *Am J Med Sci*, 1939, 188, 329
- RESNIK, H and FRIEDMAN, B *J Clin Invest*, 1935, 14, 551
- STARLING, E H and associates The regulation of the heart beat *J Physiol*, 1914, 48, 465 On the mechanical factors which determine the output of the ventricles *J Physiol*, 1914, 48, 357 Kozova, S The mechanical regulation of the heart beat in the tortoise *J Physiol*, 1915, 40, 233 *J Physiol*, 1927, 62, 243
- STARR, I and associates *J Clin Invest*, 1934, 13, 561 *Am Heart Jour*, 1943, 26, 291
- TAYLOR, N B and CAMERON, H G *Am J Physiol*, 1922, 61, 385
- VINCENT, S and THOMPSON, J H *J Physiol*, 1928, 66, 307
- VISSCHER, M B *Am Heart J*, 1938, 16, 143 *Am Ass Adv Sci*, 1940, No 13, p 176
- VISSCHER, M B and STARLING, E H *J Physiol*, 1927, 62, 243
- WARREN, J and STEAD, E A *Arch Int Med*, 1944, 73, 138

Monographs and Reviews

- ANREP, G V *Studies in cardiovascular regulation* Lane Lecture, Stanford Univ Press, 1936
- GROLLMAN, A The cardiac output of man in health and disease Thomas, Springfield, 1932
- HARRISON, T R The pathogenesis of congestive heart failure *Medicine*, 1935, 14, 255 Failure of the circulation, The Williams & Wilkins Co, Baltimore, 1936 Arterial and venous pressure factors in circulatory failure *Physiol Rev*, 1938, 18, 86
- HEYMANS, C The pressoreceptive mechanisms for the regulation of the heart rate, vasomotor tone, blood pressure and blood supply *New Eng J Med*, 1938, 219, 147
- LEWIS, T *Diseases of the heart* described for practitioners and students Macmillan, London, 1933
- LOEWI, O The humoral transmission of nervous impulses *Harvey Lectures* 1932-33, 28, 218 On problems connected with the principle of humoral

- transmission of nervous impulses Proc Roy Soc , 1935, B 118, 229
- STARLING, E H The Linacre lecture on the law of the heart Longmans, London, 1918
- STARR, I Our changing view point about congestive heart failure Ann Int Med , 1949, 30, 1

CHAPTER 26

THE OUTPUT OF THE HEART

- BARCROFT, J, BOCK, A V and ROUGHTON, F J Heart, 1922, 9, 7
- BAZETT, H C and associates Am J Physiol , 1935, 113, 312
- COURNAND, A and associates J Clin Invest , 1942, 21, 287 Proc Soc Exp Biol Med , 1941, 46, 462 J Clin Invest , 1942, 21, 287 Proc Fed Am Soc Exp Biol , 1945, 4, 207
- DOUGLAS, C G and HALDANE, J S J Physiol , 1922, 56, 69
- FORSSMANN, W I Klin Wehnschr , 1929, 8, 2085
- HAMILTON, W F and associates Am Jour Physiol , 1932, 99, 534
- HARRISON, T R, DOCK, W and HOLMAN, E Heart, 1924, 11, 337
- KEYES, A and associates Am J Roent , 1940, 44, 805
- KLEIN, O München Med Wehnschr , 1930, 77, 1311
- LEWIS, T and DRURY, A N HEART, 1923, 10, 301
- MASTER, A M and associates Arch Int Med , 1942, 69, 1010
- McMICHAEL, J and SHARPY-SCHAFER, E Br Heart J , 1944, 6, 33
- SCHNEIDER, E C and CRAMPTON, C B Am J Physiol , 1934, 110, 14
- SNYDER, J C J Clin Invest , 1938, 17, 571
- STARR, I and associates Am J Physiol , 1939, 127, 1 J Clin Invest , 1940, 19, 765
- STEWART, G N, Jour Physiol , 1894, 15, 31

Monographs and Reviews

- BAINBRIDGE, F R The physiology of muscular exercise Third edition, rewritten by BOCK, A V and DILL, D B Longmans, London, 1931
- GROLLMAN, A The cardiac output in health and disease Thomas, Springfield, 1932
- MARSHALL, E K, JR The cardiac output of man Medicine, 1930, 9, 175
- McMICHAEL, J Cardiac output in congestive heart failure Quart J Med , 1938, 7, 331

CHAPTER 27

THE CONTROL OF THE BLOOD VESSELS THE VASOMOTOR MECHANISMS PERIPHERAL VASCULAR DISORDERS SURGICAL SHOCK

- ALAM, M and SMIRK, F H J Physiol , 1937, 89, 372
- ANREP, G V and associates J Physiol , 1935, 85, 409 Ibid , 1936, 86, 431
- BACH, L M N Am J Physiol , 1946, 145, 474
- BARRON, D H and MATTHEWS, B H. C J Physiol 1935, 85, 104
- BARSOUM, G S and GADDUM, J H. Clin Sci , 1936, 2, 357
- BAYLISS, W M On the physiology of the depressor nerve J Physiol , 1893, 14, 303 On the origin from the spinal cord of the vaso-dilator fibres of the hind-limb, and on the nature of these fibres Ibid , 1901, 26, 173 Further researches on antidromic nerve-impulses Ibid , 1902, 28, 276 On reciprocal innervation of vasomotor reflexes and on the action of strychnine and chloroform thereon Proc Roy Soc , 1908, B 80, 339
- BERNTHAL, T Am J Physiol , 1938, 121, 1
- BEST, C H, DALE, H H, DUDLEY, H W and THORPE, W V J Physiol , 1927, 62, 397
- BEST, C H. and McHENRY, E W J Physiol , 1930, 70, 349
- BISHOP, G H HEINBECKER, P and O'LEARY, J L Am J Physiol , 1933, 106, 647
- BLALOCK, A and associates Series of papers in Arch Surg from 1930, Vol 20, onwards
- BOLTON, B, CARMICHAEL, E A, and STÜRUP, G J Physiol , 1936, 86, 83
- BRONK, D W and STELLA, G J Cell Comp Physiol , 1932, 1, 113 Am J Physiol , 1935, 110, 708
- BRUCE, A N Quart J Exp Physiol , 1913, 6, 339
- BÜLBRING, E and BURN, J H. J Physiol , 1936, 88, 341
- BURN, J H J Physiol , 1923, 75, 144
- CANNON, W B Ann Surg , 1934, 100, 704
- CANNON, W B FRASER, J and HOOPER, A N Med Res Com , London, Spec Rep , No 29, 1919
- CARLSON, A J Am J Physiol , 1907, 19, 408
- CODE, C F J Physiol , 1937, 90, 485
- COLLINS, W S and WILENSKY, N D J Am Med Assoc , 1936, 107, 1960
- COMROE, J H. Am J Physiol , 1938, 123, 41
- CYON, E VON and LUDWIG, C Ber Sachs Ges , 1866, 18, 307 (Quoted by Bayliss)
- DALE, H H J Physiol , 1913, 46, 291 J Pharm Exp Therap , 1912, 4, 167
- DALE, H H and EVANS, C L J Physiol , 1922, 56, 125
- DALE, H H and LAIDLAW, P P J Physiol , 1910, 41, 318 Ibid , 1911, 43, 182 Ibid , 1919, 52, 355
- DALE, H H and RICHARDS, A. N J Physiol , 1918, 52, 110
- DALY, I. DE B and VERNEY, E B J Physiol , 1927, 62, 330
- DASTRE, A and MORAT, J P Compt rend Acad Sci , 1880, 91, 393 (Quoted by Bayliss)
- DENSHAM, H B and WELLS, H M Quart J Exp Physiol , 1927, 18, 175 Ibid , 291
- ABRAMSON, D I and associates Proc Soc Exp Biol Med , 1941, 48, 214

- DOR, Y J *Physiol*, 1920, 54, 227
- DRUMMOND, H. and TAYLOR, E S *Med Res. Coun. Spec. Rep* No 25, 1919
- DRURY, A. N., FLOREY, H. and FLOREY, M E J *Physiol.*, 1929, 68, 173 See also, BARCROFT, J and FLOREY, H. The effect of exercise on the vascular conditions in the spleen and the colon. *Ibid*, p 181
- DRURY, A. N. and SZENT-GYÖRGYI, A. J *Physiol*, 1929, 68, 213
- EGGLETON, M G *Lancet*, 1944, 2, 208
- ENGLE, G L *Fainting*, Am Lecture Series C C Thomas, Springfield, Ill, 1950
- ERLANGER, J and associates *J Am Med Assoc.*, 1917, 69, 2089 *Am J Physiol*, 1919, 49, 345
- ESSEX, H. E. and associates *Am. J Physiol*, 1943, 139, 351
- FATHERBEE, T J and ALLEN, E V *Arch. Int. Med*, 1938, 62, 1015
- FELDBERG W *Histamine and anaphylaxis Ann Rev Physiol.*, 1941, 3, 671
- FOERSTER, O *Brain*, 1933 56, 1
- FREEMAN, N E *Am J Physiol*, 1933, 103, 185
- FREY, E K and KRAUT, H. Z *physiol Chem*, 1926, 167, 32
- GAMMON, G D and BRONE, D W *Proc. Soc. Exp Biol. Med*, 1934, 31, 788
- GIBBON, J H. and LANDIS, E M *J Clin. Invest.*, 1932, 11, 1019
- GRANT, R. T J *Clin Science*, 1938, 3, 157
- HEAD, H. and CAMPBELL, A. W *Brain*, 1900, 23, 353
- HELMER, O M and associates *Am. Heart J*, 1939, 17, 15
- HEREMAN, L G and RETD, M R. *Arch Surg*, 1934, 29 697
- HERRICK, J F, ESSEX, H. E. and BALDES, E J *Am. J Physiol*, 1932, 101, 213 *Ibid*, 1933, 103, 592
- HEYMANS, C. Nobel Prize Lecture, Stockholm, 1946
- HOLLINGSHEAD, W. H. *Am J Physiol*, 1947, 147, 654.
- IZQUIERDO, J J J *Physiol*, 1930, 70, 221
- KAHR, S and SHEEHAN, D *Brain*, 1933, 56, 265
- KALE, H. *Klin. Woch.*, 1929, 8, 64
- KATZ, L. N., LINDNER, E. and LANDT, H. *J Clin Invest.* 1934, 13, 37
- KEITH, N M *Med Res Com.*, London, *Spec. Rep* No 27, 1919
- KISSIN, M J *Clin Invest.*, 1934, 13, 37
- KNISELY, M H. and BLOCK, E. H *Anat. Rec.*, 1942, 82, 426 *Science*, 1947, 106, 431
- KURÉ, K and associates. *Quart. J Exp Physiol.*, 1928, 18, 333
- LA GLEY, J N and ANDERSON, H. K. *J Physiol*, 1894, 16, 410
- LEE W E and associates *Penn Med J*, 1941, 44, 1114.
- LEWIS, T Experiments relating to the peripheral mechanism involved in spasmodic arrest of the circulation in the fingers, a variety of Raynaud's disease. *Heart*, 1929, 15, 7 Raynaud's disease, with special reference to the nature of the malady *Brit. Med J*, 1932, 2, 136 Vasovagal syncope and the carotid sinus mechanism. *Ibid*, 1932, 1, 873 Clinical observations and experiments relating to burning pain in the extremities and to so-called "erythromelalgia" in particular *Clin Sci*, 1933, 1, 175
- LEWIS, T and HESS, W *Clin Sci*, 1933, 1, 39
- LEWIS, T and LANDIS, E M *Heart*, 1929, 15, 151 *Ibid*, 1930, 15, 329
- LEWIS, T and PICKERING, G W Vasodilatation in the limbs in response to warming the body, with evidence for sympathetic vasodilator nerves in man *Heart*, 1931, 16, 33 Observations upon maladies in which the blood supply to digits ceases intermittently or permanently, and upon bilateral gangrene of digits, observations relevant to so-called "Raynaud's disease" *Clin Sci*, 1934, 1, 327
- LOVE, C *Ber Sachs Ges (Leipzig)*, 1866, 18, 95 (Quoted by Bayliss)
- MAYER, J F and SOLAARDT, D Y *Am J Physiol.*, 1941, 133, 376
- MANN, F C *Surg. Gynec. Obstet.*, 1915, 21, 430
- MANWARING, W H. *Z Immun exp Therap*, 1910, 8, 1
- MARKOWITZ, J *Surg. Gyn Obstet*, 1952, 95, 644
- MARTIN, E G and STILES, P G *Am J Physiol*, 1916, 40, 194
- MATHISON, G C. *J Physiol.*, 1910, 41, 416 *Ibid*, 1911, 42, 283
- MAYERSON, H. S *Am J Physiol.*, 1942, 135, 411
- MAZUR, A and SHORR, E J *Biol. Chem*, 1948, 176, 771
- MCDOWALL, R. J S *J Physiol*, 1924, 59, 41
- MOORESON, R. A. and HOOKER, D R *Am J Physiol.*, 1915, 37, 86
- MUDD, S and GRANT, S B *J Med Res*, 1919, 40, 53
- MULINOS, M G and SHULMAN, I. *Am J Physiol*, 1939, 125, 310
- NOBLE, R. L. and TAYLOR, N B G *J Physiol.*, 1953, 122, 220
- O'SHAUGHNESSY, L. and SLOME, D Etiology of traumatic shock. *Brit. J Surg.*, 1935, 22, 589
- PARSONS, E and PREMISTER, D B *Surg, Gynec. Obstet.*, 1930, 51, 196
- PARTRIDGE, R C *Trans Roy Soc. Can*, 1932, 26, Sec. V, 129
- PAWLOV, I *Pflüger's Arch.*, 1879, 20, 210
- PICKERING, G W and WAYNE, E J *Clin. Sci*, 1934, 1, 305
- QUÉRU, E *Rev de Chir*, 1918, 56, 204 (Quoted by Cannon)
- RANSON, S W and BILLINGSLEY, P R. Vasomotor reactions from stimulation of the floor of the fourth ventricle. Studies in vasomotor reflexes arcs III.

- Am J Physiol., 1916, 41, 85 New evidence in favor of a chief vasoconstrictor center in the brain Studies in vasomotor reflexes arcs IV Ibid, 1916, 42, 1. Afferent spinal path for the depressor reflex. Studies in vasomotor reflex arcs Ibid, p 9
- RANSON, S W and HESS, C L VON Am J Physiol, 1915, 38, 128
- RHOADES, J E and associates Ann Surg, 1941, 113, 955
- ROBERTSON, B and BOYD, G L J Lab Clin Med, 1923, 9, 1
- ROSE, B and BROWNE, J S L Ann Surg, 1942, 115, 390
- ROSENBLEUTH, A and CANNON, W B Am J Physiol, 1934, 108, 599
- SIMONART, A Arch internat Pharmacodyn Therap, 1930, 37, 269
- SIMPSON, L, BROWN, G E and ADSON, A W Arch Neur Psychiat, 1931, 26, 687
- SOLLMANN, T and BROWN, E D Am J Physiol, 1912, 30, 88
- SPEISS, G Münch med Woch, 1906, 53, 345 (Quoted by Bruce)
- STOPFORD, J S B and TELFORD, E D Brit J Surg 1931, 18, 557
- STRICKER, S Wien Sitzber, 1876, 74, (III) 173 (Quoted by Bayliss)
- SWINGLE, W W and associates Science, 1933, 77, 58
- TAYLOR, N B and MOORHOUSE, M S Nat Res Coun Report, Ottawa, 1941
- TAYLOR and NOBLE see NOBLE and TAYLOR
- TODD, T W and KRAMER, J G Anat Rec, 1914, 8, 243
- TRUETA, J Lancet, 1946, II, 237 See also under Reviews and Monographs
- UPRUS, V, GAYLOR, J B and CARMICHAEL, E A Clin Sci, 1936, 2, 301
- VINCENT, S and THOMPSON, J H J Physiol, 1928, 66, 307
- VOGT, E Z exp Path therap, 1912, 11, 191
- WANG, S C and BORISON, H L Am J Physiol, 1947, 150, 712, 722
- WATERS, E T and associates J Physiol, 1941, 99, 454
- WEISS, S WILKINS, R W and HAYNES, F W J Clin Invest, 1937, 16, 73 Ibid, 17, 41
- WILKINS, R W and EICHNA, L W Bull Johns Hopk Hosp, 1941, 68, 425
- WILSON, W C and associates Lancet, 1936, 1, 1400
- ZUCKERMANN, S Proc Roy Soc. Med, 1941, 34, 171,
- ZWEIFACH, B W and associates Am J Physiol, 1944, 142, 80 Ann Surg, 1944, 120, 232 Surg Gynec. Obstet, 1945, 80, 593 Ann N Y Acad Sc, 1946 46, 553, 571 Am J Physiol, 1947, 150, 239, Shock and Circulatory Homeostasis Trans 1st Conf, J Macy Found, 1951, p 15
- Monographs and Reviews*
- ANREP, G V The circulation in striated and plain muscles in relation to their action Harvey Lectures, 1934-35, 30, 146 Studies in cardiovascular regulation Lane Lectures, Stanford Univ Press, 1936
- BAYLISS, W M The vaso-motor system Longmans, London, 1923
- BEST, C H and MCHENRY, E W Histamine Physiol Rev, 1931, 11, 371
- BRONK, D W The nervous mechanism of cardiovascular control Harvey Lectures, 1933-34, 29, 245
- BROWN, G E and ALLEN, E V Thrombo-angitis obliterans, clinical, physiologic and pathologic studies Saunders, Philadelphia, 1928
- BURN, J H Sympathetic vasodilator fibres Physiol Rev, 1938, 18, 137
- DRAGSTEDT, C A. Anaphylaxis Physiol Rev, 1941, 21, 563
- FELDBERG, W Histamine and anaphylaxis Ann Rev Physiol, 1941, 3, 671
- FRANKLIN, K J The physiology and pharmacology of veins Physiol Rev, 1928, 8, 346
- FREEMAN, N E Burns, shock, wound healing and vascular injuries Saunders, Phila, 1943
- GREEN, N D Shock Anesthesiology, 1942, 3, 611
- GREGG, M I Physiological contributions to the problem of shock Fed Proc, 1946, 5, 355
- HARKINS, H N Recent advances in the study and management of traumatic shock Surgery, 1941, 9, 231, 447, 607 The present status of the problem of thermal burns Physiol Rev, 1945, 25, 531
- HEYMANS, C The pressoreceptive mechanisms for the regulation of heart rate, vasomotor tone, blood pressure and blood supply New Eng J Med, 1938, 219, 147
- HEYMANS, C BOUCKAERT, J J and REGNIER, P La sinus carotidien et la zone homologue cardioarotique Doin, Paris, 1933
- LEWIS, T The blood vessels of the human skin and their responses Shaw, London, 1927
- LEWIS, T Clinical science, illustrated by personal experiences Shaw, London, 1934 Vascular disorders of the limbs Macmillan, London, 1936
- LEWIS, T Observations on some normal and injurious effects of cold upon the skin and underlying tissues Brit Med J, 1941, II, 795, 837, 869
- MANN, F C and ESSEX, H E The present status of the problem of traumatic shock Am J Surg, 1935, NS 28, 161
- MARKOWITZ, J Surg Gynec Obst, 1952, 95, 644
- MCDOWALL, R J S The nervous control of the blood vessels Physiol Rev, 1935, 15, 98
- RANSON, S W Afferent paths of visceral reflexes Physiol Rev, 1921, 1, 477
- RICHARDS, R L The peripheral circulation in health and disease Livingstone, Edinb, 1946

- SCUDDER, J Shock, blood studies as a guide to therapy Lippincott, Philadelphia, 1940
- Traumatic Shock Discussion, Proc. Roy Soc Med, London, 1935, 28, 1473
- WALLACE, C S, DALE, H H, BAYLISS, W M, CANNON, W B and others Traumatic toxæmia as a factor in shock Med Res Com, London, Spec Rep No 26, 1919
- WEISS, S and BAKER, J P The carotid sinus reflex in health and disease Medicine, 1933, 12, 297 See also, FERRIS, E B, CAPPS, R B and WEISS, S Carotid sinus syncope and its bearing on the mechanism of the unconscious state and convulsions Ibid, 1935, 14, 457
- WIGGERS, C J The present status of the shock problem Physiol Rev, 1942, 22, 74
- ZUCKERMAN, S Proc. Roy Soc. Med, 1941, 34, 171
- CHAPTER 28
- SPECIAL FEATURES OF THE CIRCULATION IN DIFFERENT REGIONS
- ANREP, G V and associates Brit Heart J, 1946, 8, 171
- ANREP, G V and BULATAO, E Observations on the pulmonary circulation Pulmonary circulation in the heart-lung preparation J Physiol, 1925, 60, 175
- ANREP, G V, CRUIKSHANK, E W H, DOWNING, A C and SUBBA RAU, A The coronary circulation in relation to the cardiac cycle Heart, 1927, 14, 111
- ANREP, G V and HAUSLER, H The coronary circulation I The effect of changes of the blood-pressure and of the output of the heart J Physiol, 1928, 65, 357 ANREP, G V and KING, B The significance of the diastolic and systolic blood pressures for the maintenance of the coronary circulation Ibid, 64, 341 ANREP, G V and SEGALL, H N The regulation of the coronary circulation Heart, 1926, 13, 239 ANREP, G V and STACEY, R S Comparative effect of various drugs upon the coronary circulation J Physiol, 1927, 64, 187 See also, HAUSLER, H The coronary circulation III. The dependence of changes in the coronary blood flow on cardiac and local vascular factors Ibid, 1929, 68, 324 MARKWALDER, J and STARLING, E H A note on some factors which determine the blood flow through the coronary circulation Ibid, 1913, 47, 275
- BALDES, E J and HERRICK, J F Proc Soc Exp Biol Med, 1937, 37, 43
- BARCROFT, J, NISIMARU, Y and RAY, G B J Physiol, 1932, 74, 44
- BARSOUM, G S and SMIRK, F H. Clin Sci, 1936, 2, 337
- BAUER, W, DALE, H H, POULSSON, L T and RICHARDS, D W J Physiol, 1932, 74, 343
- BECK, C S and TICHY, V L Am Heart J, 1935, 10, 849
- BORDLEY, J, GROW, M H and SHERMAN, W B Bull J Hopk Hosp, 1938, 62, 1
- BRADFORD, J R and DEAN, H. P Proc Roy Soc, 1889, B 45, 369 J Physiol, 1894, 16, 34
- BRADLEY, S E and associates Jour Clin Invest, 1945, 24, 890
- CHAMBERS, R and ZWEIFACH, B W Topography and function of the mesenteric capillary circulation. Am J Anat, 1944, 75, 173 Intercellular cement and capillary permeability Physiol Rev, 1947, 27, 436 See also ZWEIFACH, B W The character and distribution of the blood capillaries Anat Rec 1939, 74, 475 The structural basis of permeability and other functions of blood capillaries Cold Spring Harbor Symposia on Quant Biol, 1940, 8, 216
- CLARK, E R and CLARK, E L Am J Anat, 1925, 35, 239, Ibid, p 265 Caliber changes in the minute vessels observed in the living animals Am J Anat., 1943, 73, 215
- COBB, S Assoc Res Nerv Ment Dis, 1929, 9, 458
- COTTON, T F, SLADE, J G and LEWIS, T Heart, 1917, 6, 227
- DALE, H Lancet, 1929, 1, 1179, 1233, 1285
- DALE, H. and RICHARDS, A. N J Physiol, 1918, 52, 110
- DALY, I DEB and EULER, U S VON Proc Roy Soc., 1932, B 110, 92
- DALY, I. DEB and associates Quart J Exp Physiol, 1937, 27, 123
- DOI, Y J Physiol, 1920, 54, 227
- DURAN REYNALS, F Bact Rev, 1942, 6, 197 See also McCLEAN, D, Jour Path and Bact, 1933, 33, 1045 Ibid, 1936, 44, 477, see also CHAIN, E and DUTHIE, Nature, 1939, 144, 977, SANNELLA, L S Yale Jour Biol and Med 1940, 12, 433, HECHTER, O and associates Journ Pediat, 1947, 30, 645
- ECKENHOFF, J E and associates Am J Physiol, 1947, 148, 582
- ESSEX, H. E Am Ass Adv Sci, 1940, 13, 61
- ESSEX, H. E and associates Am J Physiol Proc, 1939, 126, 486 Influence of exercise on blood pressure, pulse rate and coronary flow of the dog Am J Physiol, 1939, 125, 614 Control of coronary blood flow in the heart lung preparation Ibid, 1940, 131, 43
- FELDBERG, W and KRAYER, O Arch Exp Path Pharmacol, 1933, 172, 170
- FERRIS, E B Arch Neur Psych, 1941, 46, 377
- FINESINGER, J E and PUTNAM, T J Arch Neur Psychiat, 1933, 30, 775
- FINESINGER, J E and COBB, S Arch Neur Psychiat, 1933, 30, 980
- FIROR, W M and STINSON, E, JR. Bull Johns Hopkins Hosp, 1929, 44, 138
- FLOREY, H W and CARLETON, H M Proc Roy Soc, 1926, B 100, 23
- FORBES, H S Arch Neur Psychiat, 1928, 18, 751

- FORBES, H S and WOLFF, H G The vasomotor control of cerebral vessels *Arch Neur Psychiat*, 1928, 19, 1057 Observations of the pial vessels during changes in intracranial pressure *Assoc Res Nerv Ment Dis*, 1929, 8, 91 See also, LENNOX, W G The cerebral circulation. The respiratory quotient of the brain and of the extremities in man *Arch Neur Psychiat*, 1931, 26, 719 The effect of mental work *Ibid*, p 725 TALBOTT, H. H., WOLFF, H G and COBB, S Changes in cerebral capillary bed following cervical sympathectomy *Ibid*, 1929, 21, 1102
- FORBES, H. S., NASON, G I. and WORTMAN, R C Vasodilation in the pia following stimulation of the vagus, aortic and sinus nerves *Arch Neur Psychiat.*, 1937, 37, 334
- GIBBS, F A *Arch. Neur Psychiat*, 1933, 30, 1003
- GRANT, R T *Heart*, 1930, 15, 527
- GRANT, R T and VIKO, L E *Heart*, 1929, 15, 103
- GREGG, D E *Am Ass Adv Sci*, 1940, 13, 81
- GREGG, D E and associates *Am J Physiol*, 1943, 139, 732
- GREGG, D E and SHIPLEY, R E *Am Jour Physiol*, 1947, 151, 13
- GROSS, L and KUGEL, M A. *Am Heart J*, 1933, 9, 165
- GULLAND, L *Brit Med J*, 1898, 2, 781
- HEAD, H. and CAMPBELL, A W *Brain*, 1900, 23, 353
- HEINBECKER, P *Am J Physiol*, 1927, 81, 170, HEINBECKER, P and BARTON, W A. *Ann Surg*, 1941, 114, 186
- HEPBURN, J and JAMIESON, R A *Am Heart J*, 1926, 1, 623
- HEYMANS, C and BOUCKAERT, J J *J Physiol*, 1933, 79, 49 *Ibid*, 1935, 84, 367
- HILTON, R and EICHHOLTZ, F J *Physiol*, 1925, 59, 413
- HOOKE, D R *Am J Physiol*, 1920, 54, 30
- HUBER, G C *J Comp Neur*, 1899, 9, 1
- HUDSON, C L, MORITZ, A R and WEARN, J T *J Exp Med*, 1932, 56, 919
- KABAT, H and associates (ROSSEN, KABAT and ANDERSON) *Arch Neurol Psychol*, 1943, 50, 510
- KATZ, L N, HAMBURGER, W W and SCHUTZ, W J *Am Heart J*, 1934, 9, 771
- KATZ, L N and LINDNER, E *Am J Physiol*, 1939, 126, 283
- KATZ, L N and associates *Am J Physiol*, 1933, 106, 183
- KROGH, A I The reaction to stimuli and the innervation of blood vessels in the tongue of the frog *J Physiol*, 1920, 53, 399 II The reactions to local stimuli of the blood vessels in the skin and web of the frog *Ibid*, 1921, 55, 412 KROGH, A, HARROP, G A and REHBERG, P B III The innervation of blood vessels in the hind limb of the frog *Ibid*, 1922, 56, 179
- LANDIS, E *Heart*, 1930, 15, 209 *Ibid*, p 247
- LEWIS, T *Heart*, 1924, 11, 151
- LEWIS, T and GRANT, R T *Heart*, 1925-1926, 12, 73
- LEWIS, T and HARMER, I M *Heart*, 1927, 14, 19
- LOMBARD, W P *Am J Physiol*, 1912, 29, 335
- MANN, F C and MAGATH, *Am J Med Sci*, 1921, 161, 37 *Am J Physiol*, 1921, 55, 285
- MANN, F C and associates The blood flow in the left coronary artery of the dog *Am J Physiol*, 1935, 113, 39 (*Proc*)
- MARKOWITZ, J and associates *Am J Dig Diseases*, Oct, 1949, p 344
- MARKOWITZ, J and SOSKIN, S *Proc Soc Exp Biol Med*, 1927, 25, 7 SOSKIN, S *J Lab Clin Med*, 1931, 16, 382
- MARVIN, H. M *Bull N Y Acad Med*, 1941, 17, 737
- MARVIN, H M and HARRIS, K E *Heart*, 1927, 14, 135
- MAUTZ, F R and GREGG, D E *Proc Soc Exp Biol Med*, 1937, 36, 797 *Am J Physiol*, 1939, 127, 161
- MOE, G K and VISSCHER, M B *Am Ass Adv Sci*, 1940, 13, 100
- O'SHAUGHNESSY, L *Lancet*, 1937, 1, 185
- PARDEE, H. E B *Arch Int Med*, 1920, 26, 244
- PARKINSON, J and BEDFORD, D E *Heart*, 1928, 14, 195
- PENFIELD, W Intracerebral vascular nerves *Arch Neur Psychiat*, 1932, 27, 230
- PICKERING, G W and WAYNE, E J Observations on angina pectoris and intermittent claudication in animals *Clin Sci*, 1934, 1, 305 See also, WAYNE, E J and GRAYBIEL, A Observations on the effect of food, gastric distension and repeated exercise on angina of effort, with a note on angina sine dolore *Ibid*, p 287 WAYNE, E J and LAPLACE, L B Observations on angina of effort *Ibid*, p 103
- PORTER, W T *Am J Physiol*, 1898, 1, 145
- PRATT, F H *Am J Physiol*, 1898, 1, 86 See also, BATSON, O V and BELLET, S *Am Heart J*, 1930, 6, 206
- ROUS, P and GILDING, H P *Proc Soc Exp Biol and Med*, 1929, 26, 497
- ROY, C S and BROWN, J G *J Physiol*, 1880, 2, 323
- SCARAMUCCI *Diario Parmense* 1689 Quoted by Porter from Haller's *Elementa Physiolog*, 1778, lib iv, 459
- SCHLESINGER, M J *Am Ass Adv Sci*, 1940, 13, 61
- SCHMIDT, C F *Proc Fed Am Soc Exp Biol*, 1944, 3, 131
- SCHWIEGE, H. *Pflüger's Arch*, 1935, 236, 206
- SMITH, F M *Arch Int Med*, 1918, 22, 8 *Ibid*, 1923, 32, 497
- STELLA, G *J Physiol*, 1931, 73, 36
- STROEM, Nerva theoria machinae animalis (Quoted by Porter from Haller)
- SULTZER, R *Heart*, 1924, 11, 141
- SUTTON, D C and LUETH, H C *Proc Soc Exp Biol Med*, 1928, 25, 842
- THEBESTIUS, A C *Dissert Med Inaug de circulo*

- sanguinis in corde. Leiden, 1708 (Quoted by P D White.)
- VIETSSSENS, R. Nouvelle découverte sur le cœur TOULOUSE, 1706 (Quoted by Wearn.)
- WEARN, J T The extent of the capillary bed of the heart. *J Exp Med*, 1928, 47, 273 The rôle of the Thebesian vessels in the circulation of the heart. *Ibid*, p 293
- WEARN, J T, SHIPLEY, R. A. and SHIPLEY, L J The capillary supply in normal and hypertrophied hearts of rabbits *Ibid*, 1937, 65, 29
- WEISS, S and LENOX, W G *Arch. Neur Psychiat.*, 1931, 26, 737 See also, FORBES, H. S., WOLFF, H. G and COBB, S *Am J Physiol.*, 1929, 89, 266
- WIGGERS, C. J The action of adrenalin on the pulmonary vessels *J Pharm. Exp Therap*, 1909, 1, 241 The innervation of the coronary vessels *Am J Physiol*, 1909, 24, 391 On the action of adrenalin on the cerebral vessels *Ibid*, 1905, 14, 452
- WOOLLARD, H. H. The innervation of blood vessels *Heart*, 1926, 13, 319 The innervation of the heart. *J Anat*, 1926 60, 345
- Monoographs and Reviews*
- ANREP, G V The regulation of the coronary circulation. *Physiol Rev*, 1926 6, 596 The circulation in striated and plain muscles in relation to their activity *Harvey Lectures*, 1934-35, 30, 146 *Studies in cardiovascular regulation. Lane Lectures*, Stanford Univ Press, 1936
- CHAMBERS, R. and ZWEIFACH, B W Inter cellular cement and capillary permeability *Physiol. Rev*, 1947, 27, 436
- DALY, I DEB Reactions of the pulmonary and bronchial blood vessels *Physiol. Rev*, 1933, 13, 149 The physiology of the bronchial vascular system *Harvey Lectures*, 1935-6, p 235
- DANIELLI, J F and STOCK, A. The structure and permeability of capillaries *Biol. Rev*, 1944, 19, 81
- GREGG, D E The coronari circulation *Physiol Rev*, 1946, 26, 28
- GROSS, L. The blood supply to the heart, its anatomical and clinical aspects Hoeber, New York, 1921
- HERRICK, J B The coronary artery in health and disease. *Harvey Lectures*, 1930-31, 26, 129
- HILL, L. The physiology and pathology of the cerebral circulation. Churchill, London, 1896
- HILL, L. The cerebral circulation In Schäfer's *Text-book of Physiology*, Pentland, Edinburgh, 1900, 2, 1
- HOOKE, D R. Evidence of fundamental activity on the part of the capillaries and venules *Physiol. Rev*, 1921, 1, 112
- KROGH, A The anatomy and physiology of the capillaries Yale University Press. 1929
- LANDIS, E Capillary pressure and capillary permeability *Physiol. Rev*, 1934, 14, 404. The passage of fluid through the capillary wall. *Harvey Lecture*, 1936-37, pp 70-91
- LEVINE, S A. Coronary thrombosis its various clinical features *Medicine*, 1929, 8, 245
- LEWIS, T The blood vessels of the human skin and their responses Shaw, London, 1927
- LEWIS, T *Clinical Science*, illustrated by personal experiences Shaw, London, 1934
- MANLY, F C The circulation of the liver *Quart Bull. Ind Univ Med Center*, 1942, 4, 43
- MARKOWITZ, J and assoc. *Proc. Soc. Exp Biol Med*, 1949, 70, 304, *Physiol. Rev* 1951, 31, 188
- MEYER, K. The biological significance of hyaluronic acid and hyaluronidase. *Physiol. Rev*, 1947, 27, 335
- SNYDER, C. D Recent advances in knowledge of the liver *Physiol. Rev*, 1942, 22, 54
- SUTTON, D C. and LUETH, H. Diseases of the coronary arteries Hoeber, New York, 1921
- WAKIM, K G The effect of certain substances on the intrahepatic circulation of blood in the intact animal. *Am. Heart Jour*, 1944, 27, 280
- WIGGERS, C J The regulation of the pulmonary circulation *Physiol. Rev*, 1921, 1, 239
- WOLFF, H. G The cerebral circulation. *Physiol. Rev*, 1936, 16, 545

SECTION III

CHAPTER 29

THE MECHANICS OF RESPIRATION

- ARCHIBALD, E and BROOK, A. *Am Rev Tuberc.*, 1927, 16, 111
- BACROFT, J and KARLONEN, M J *J Physiol*, 1948, 107, 153
- BROOK, D W and FERGUSON, L K. *Am. J Physiol.*, 1935, 110, 700
- BURNS, D J *Physiol.*, June 15, 1940, 26P
- CHRISTIE, R. V and MCINTOSH, C A *J Clin Invest.*, 1934, 13, 279
- DAVIS, M E and POTTER, E L *J Am Med Assoc.*, 1946, 131, 1194
- ELLIS, M J *Physiol*, 1936, 87, 298
- EVE, F C *Lancet*, 1932, 2, 995, *Clinical J*, 1943, 72, 172
- HENDERSON, Y Resuscitation. *J Am. Med. Assoc.*, 1924, 83, 758 The prevention and treatment of asphyxia in the new born *Ibid*, 1928, 90, 583 Observations on resuscitation from asphyxia and prevention and treatment of secondary pneumonia by inhalation of CO₂ *Brit. Med J*, 1931, 2, 687 Reasons for the use of carbon dioxide with oxygen in the treatment of pneumonia. *New Eng J Med*, 1932, 206, 151 HENDERSON, Y *Adventures in respiration The Williams & Wilkins Co*, Baltimore, 1938

- JOSSELYN, L E *Anat Rec*, 1935, 62, 147
 MACKLIN, C C The dynamic bronchial tree *Am Rev Tuberc*, 1932, 25, 363 Pulmonic interlobular air passages *Trans Roy Soc. Can*, Sec V, 1934, 28, 37 *J Anat.*, 1935, 69, 188
 MOLL, H H *Quart J Med*, 1932, N S 1, 457
 NEGOVSKI, V A. *Am Rev Soviet Med*, 1945, 2, 303, 408, 491 *Ibid*, 3, 45, 147, 243, 339
 NICHOLSON, T R and TRIMBY, R H. *Am J Physiol*, 1940, 128, 276
 ROBERTS, J C and BLAIR, L G *Lancet*, 1950, 1, 386
 ROBERTSON, O H Phagocytosis of foreign material in the lung *Physiol Rev*, 1941, 21, 121
 ROSS, B O *J Am Med Assoc*, 1943, 122, 660
 SHAW, L A and DRINKER, P J *Clin Invest*, 1929, 7, 229 *Ibid*, 1930, 8, 33
 SNYDER, F F and ROSENFELD, M *Am J Physiol*, 1937, 119, 153, *Anat Rec.*, 1937, 67, Supp 3, p 47
 THORNTON, J W *Quart J Exp Physiol*, 1931, 21, 305
 VAN ALLEN, C M LINDSKOG, G E and RICHTER, H G *Yale J Biol Med*, 1930, 2, 297 *J Clin Invest*, 1931, 10, 559 LINDSKOG, G E and VAN ALLEN, C M *Arch Surg*, 1932, 24, 204
 WARNER, W P *Quart J Med*, 1934, N S 3, 401
 WARNER, W P and GRAHAM, D *Arch Int Med*, 1933, 52, 888
 WEST, S *Brit Med J*, 1887, 1, 393
 WHITWELL, F *Thorax* 1952, 7, 213
 WILSON, H. G *Am J Anat*, 1922, 30, 267
 WINDLE, W F *Physiology of the foetus* W B Saunders Co, Phila, 1940 *Harvey Lect*, 1944-45, 40, 236

Monographs and Reviews

- BARCROFT, J Fetal circulation and respiration *Physiol Rev*, 1936, 16, 103
 GRAHAM, E A Alterations of intrapleural pressure and their significance *Harvey Lectures*, 1923-24, 19, 123 *Medicine*, 1924, 3, 417
 KEITH, A The mechanism of respiration in man Further advances in physiology Arnold, London, 1909
 MACKLIN, C C The musculature of the bronchi and lungs *Physiol Rev*, 1929, 9, 1
 MILLER, W S *The lung* Thomas, Springfield, 1937

CHAPTER 30

THE AIR OF THE LUNGS

- CHRISTIE R V *J Clin Invest*, 1932, 11, 1099
 COMROE, J H and DRIPPS, R D *Am J Physiol*, 1944, 142, 700
 DREYER, G *Lancet*, 1919, 2, 227
 HENDERSON, Y, CHILLINGWORTH, F P and WHITNEY, J L *Am J Physiol*, 1915, 38, 1

- HUTCHINSON, J In *Todd's Cyclopaedia of Anatomy and Physiology*, 1850, 4, p 1016
 LILIENTHAL, J L JR and associates *Am J Physiol*, 1946, 147, 199
 VAN SLYKE, D D and BINGER, C A S *J Exp Med*, 1923, 37, 457
 WEST, H. F *Arch. Int. Med*, 1920, 25, 306

Monographs and Reviews

- ARNETT, J H Vital capacity of the lungs, changes occurring in health and disease *J Clin Invest*, 1935, 14, 543
 HALDANE, J S and PRIESTLEY, J G *Respiration* Clarendon Press, Oxford, 1935
 HENDERSON, Y *Adventures in respiration* The Williams & Wilkins Co, Baltimore, 1938
 PEABODY, F W *Cardiac dyspnea* *Harvey Lectures*, 1916-17, 12, 248

CHAPTER 31

THE PHYSICAL PRINCIPLES GOVERNING THE RESPIRATORY EXCHANGES

- BARCROFT, J and NAGAHASHI, M *J Physiol*, 1921, 55, 339
 BOCK, A V and associates *J Physiol*, 1929, 68, 277
 CHRISTIE, R V and LOOMIS, A L *J Physiol*, 1932, 77, 35
 GROOM, D and associates *Proc Staff Meet Mayo Clin*, 1948, 23, 601
 KROGH, A Some new methods for the tonometer determination of gas-tension in fluids *Skand Arch Physiol*, 1908, 20, 259 On micro-analysis of gases *Ibid*, p 279
 KROGH, A and KROGH, M *Skand Arch. Physiol*, 1910, 23, 179
 LILIENTHAL, J L and associates *Am J Physiol*, 1946, 147, 199
 MILLIKIN, G A *Rev Scientif Instr*, 1942, 13, 434
 ROUGHTON, F J W *Am J Physiol*, 1945, 143, 594
 SQUIRE, J R *Clin Sci*, 1940, 4, 33

Monographs and Reviews

- BAINBRIDGE, F A *The physiology of muscular exercise*, 3rd ed, rewritten by BOCK, A. V and DILL, D B Longmans, London, 1931
 PETERS, J P and VAN SLYKE, D D *Quantitative clinical chemistry* Vol II The Williams & Wilkins Co, Baltimore, 1932

CHAPTER 32

- THE TRANSPORT AND DELIVERY OF OXYGEN TO THE TISSUES, INTRACELLULAR OXIDATIONS AND ENERGY TRANSFER, THE CARRIAGE OF CARBON DIOXIDE
 BARRON, E S G and SINGER, T P *J Biol Chem*, 1945, 257, 221, 241
 CHRISTIANSEN, J, DOUGLAS, C G and HALDANE, J S *J Physiol*, 1914, 48, 244

- CROSS, R. J., TAGGART, J. V., COVO, G. A., and GREEN, D. E. *J Biol Chem*, 1949, 177, 655
- DAVENPORT, H. W. Carbonic anhydrase in tissues other than blood *Physiol Rev*, 1946, 26, 574
- FERGUSON, J. K. W. and ROUGHTON, F. J. W. The direct chemical estimation of carbamino compounds of CO₂ with haemoglobin *J Physiol*, 1934, 83, 68. The chemical relationships and physiological importance of carbamino compounds of CO₂ with haemoglobin. *Ibid*, p. 87
- GORANSON, E. S. *Federation Proceedings*, 1949, 8, 58
- GRAFFLIN, A. A., and GREEN, D. E. *J Biol Chem*, 1948, 176, 95
- HAAS, E. J. *Biol Chem*, 1943, 148, 481
- HAAS, E. J. *Biol Chem*, 1944, 152, 695
- HOGEBOM, G. H. J. *Biol Chem*, 1949, 177, 847
- HORECKER, B. L. *Federation Proceedings*, 1949, 8, 207
- HORECKER, B. L. and KORNBERG, A. *J Biol Chem*, 1946, 165, 11
- JOFFE, J. and POULTON, E. P. *J Physiol*, 1920, 54, 129
- KEILIN, D. *Proc. Roy. Soc., London, B*, 1930, 106, 418
- KEILIN, D. and HARTREE, E. F. *Proc. Roy. Soc. B*, 1939, 127, 167, *Biochem J*, 1945, 39, 293, *ibid J*, 1947, 41, 500, *ibid J* 1949, 44, 205
- KEILIN, D. and MANN, T. *Nature*, 1939, 144, 442
- KENNEDY, E. P. and LEHNINGER, A. L. *J Biol Chem*, 1949, 179, 957
- LARDY, H. A. *et al Arch. Biochem*, 1945, 6, 41
- LEHNINGER, A. L. *J Biol Chem*, 1949, 178, 625
- LEHNINGER, A. L. and KENNEDY, E. P. *J Biol Chem*, 1948, 173, 753
- LE PAGE, G. A. and UMBREIT, W. W. *J Biol Chem*, 1943, 147, 263
- MEYERHOF, O. and GREEN, H. J. *Biol Chem*, 1949, 178, 655
- MEYERHOF, O. and OESPER, P. *Federation Proceedings*, 1948, 7, 174.
- OCHOA, S. *J Biol Chem*, 1943, 151, 493
- POTTER, V. R. *J Biol Chem*, 1947, 169, 17
- SLATER, E. C. *Biochem J*, 1949, 44, 305
- THEORELL, H. and ÅKESSON, Å., *J Amer Chem Soc*, 1941, 63, 1804.
- VOGLER, K. G. and UMBREIT, W. W. *J Gen. Physiol*, 1942, 26, 157
- WARNIO, W. W., COOPERSTEIN, S. J., KOLLEN, S. and EICHEL, B. *J Biol Chem*, 1948, 173, 145
- cellular respiration *Biological Symposia*, 1943, 10, 27
- BEESIN, T. *Thiolverbindungen und Enzyme. Ergeb Enzymforsch.*, 1935, 4, 68
- BREUSCH, F. L. The biochemistry of fatty acid catabolism *Advances in Enzymology*, 1948, 8, 343
- BUCHANAN, J. M. and HASTINGS, A. B. The use of isotopically marked carbon in the study of intermediary metabolism, *Physiol Rev*, 1946, 26, 120
- BURK, D. The Pasteur reaction *Cold Spring Harbor Symposia on Quantitative Biology*, 1939, 7, 420
- DIXON, K. C. The Pasteur effect. *Biol Rev. Cambridge Phil Soc.*, 1937, 12, 431
- DORFMAN, A. Pathways of glycolysis *Physiol Rev*, 1943, 23, 124
- DRABKIN, D. L. Distribution and metabolic aspects of derivatives of iron protoporphyrin *Federation Proceedings*, 1948, 7, 483
- ELLIOTT, K. A. C. Biological oxidations and reductions *Ann. Rev. Biochem*, 1946, 15, 1
- ENGELHARDT, V. A. Adenosinetriphosphatase properties of myosin. *Advances in Enzymology*, 1946, 6, 147
- EVANS, E. A. Metabolic cycles and decarboxylation *Symposium on Respiratory Enzymes*, p. 107, Univ Wisconsin Press, 1942
- FISCHER, F. G. Niedermolekulare Überträger biologischer Oxydo-Reduktionen und ihre Potentiale. *Ergeb Enzymforsch*, 1939, 8, 186
- GRANICK, S. and GILDER, H. Distribution, structure and properties of the tetrapyrroles *Advances in Enzymology*, 1947, 7, 265
- GREEN, D. E. *Mechanism of Biological Oxidations* Cambridge Univ Press, 1940
- HELLERMAN, L. Reversible inactivations of certain hydrolytic enzymes. *Physiol Rev*, 1937, 17, 454
- JAMES, W. O. The respiration of plants *Ann. Rev. Biochem*, 1946, 15, 417
- JOHNSON, M. J. The role of aerobic phosphorylation in the Pasteur effect. *Science*, 1941, 94, 200
- KALCKAR, H. M. The nature of energetic coupling in biological synthesis *Chem. Rev*, 1941, 28, 71
- KALCKAR, H. M. The function of phosphate in cellular assimilations. *Biol. Rev. Cambridge Phil Soc.*, 1942, 17, 28
- KALCKAR, H. M. The function of phosphate in enzymatic synthesis *Ann. New York Acad. Sci.*, 1944, 45, 395
- KEILIN, D. Cytochrome and intracellular respiration. *Ergeb Enzymforsch.*, 1933, 2, 239
- KREBS, H. A. The intermediary stages in the biological oxidation of carbohydrate. *Advances in Enzymology*, 1943, 3, 191
- LILLIE, R. S. *General Biology and Philosophy of Organism* University of Chicago Press, 1945
- LIPMAN, F. Metabolic generation and utilization of phosphate bond energy *Advances in Enzymology*, 1941, 1, 100

Monographs and Reviews

- BAILEY, K. The proteins of skeletal muscle. *Adv in Protein Chem*, 1944, 1, 289
- BALL, E. G. Energy relationships of the oxidative enzymes. *Ann. New York Acad. Sci.*, 1944, 45, 363
- BARROU, E. S. G. Mechanisms of carbohydrate metabolism *Advances in Enzymology*, 1943, 3, 149
- BARROU, E. S. G. The application of biological oxidation reduction reaction systems to the study of

- LIPMANN, F Pasteur Effect, Symposium on Respiratory Enzymes, Univ Wisconsin Press, 1942
- LIPMANN, F Acetyl Phosphate Advances in Enzymology, 1946, 6, 231
- LIPMANN, F Metabolic process patterns, in Currents in Biochemical Research (Editor D E Green), Interscience Pub Inc, New York, 1946
- LOEB, LEO The Biological Basis of Individuality C C Thomas, Springfield, Ill, 1945
- MEYERHOF, O Intermediate carbohydrate metabolism Symposium on Respiratory Enzymes, Univ Wisconsin Press, 1942
- MEYERHOF, O Energy relationships in glycolysis and phosphorylation Ann New York Acad Sci, 1944, 45, 377
- MICHAELIS, L Fundamentals of oxidation and reduction, in Currents in Biochemical Research (Editor D E Green), Interscience Pub Inc, New York, 1946
- OCHOA, S Cocarboxylase Biological Action of the Vitamins (Editor E A. Evans), Univ Chicago Press, 1942
- OGSTON, A G and SMITHIES, O Some thermodynamic and kinetic aspects of metabolic phosphorylation Physiol Rev, 1948, 28, 283
- OPPENHEIMER, C and STERN, K G Biological Oxidation, Nordemann, New York, 1939
- POTTER, V R Transport of hydrogen in animal tissues Medicine, 1940, 19, 441
- POTTER, V R Biological energy transformations and the cancer problem Advances in Enzymology, 1944, 4, 201
- SCHLENK, F Nicotinamide nucleotide enzymes Symposium on Respiratory Enzymes, Univ Wisconsin Press, 1942
- SCHRÖDINGER, E What is Life? The MacMillan Co, New York, 1945
- SHEN, S C The interaction of myosin and ATP Ann New York Acad Sci, 1947, 47, 875 (conference on muscular contraction)
- STADIE, W C The intermediary metabolism of the fatty acids Physiol Rev, 1945, 25, 395
- STILES, W Plant Respiration (No 2) Botanical Rev 1946, 12, 165
- STOTZ, E Pyruvate Metabolism Advances in Enzymology, 1945, 5, 129
- SZENT-GYÖRGYI, A Chemistry of Muscular Contraction Academic Press Inc, New York, 1947
- SZENT-GYÖRGYI, A. Nature of Life, a Study on Muscle Academic Press Inc, New York, 1948
- THEORELL, H Heme-linked groups and mode of action of some hemoproteins Advances in Enzymology, 1947, 7, 265
- WARBURG, O The respiration enzyme Science, 1925, 61, 57, *ibid*, 1928, 68, 437
- WEIL-MALHERBE, H. Biological oxidations and reductions Ann Rev Biochem, 1948, 17, 1
- WIELAND, H On the Mechanism of Oxidation Yale Univ Press, 1932
- WOOD, H G The fixation of carbon dioxide and the inter-relations of the tricarboxylic acid cycle Physiol Rev, 1946, 26, 198

Monographs and Reviews

- HALDANE, J S and PRIESTLEY, J G Respiration Clarendon Press, Oxford, 1935
- HENDERSON, L J Blood A study in general physiology Yale University Press, 1928
- ROUGHTON, F J W Recent work on carbon dioxide transport by the blood Physiol Rev, 1935, 15, 241
- VAN SLIKE, D D The carbon dioxide carriers of the blood Physiol Rev, 1921, 1, 141

CHAPTER 33

THE CONTROL OF RESPIRATION, PERIODIC RESPIRATION, DYSPNEA

- ADRIAN, E D J Physiol, 1933, 79, 332
- ADRIAN, E D and BRONK, D W J Physiol, 1928, 66, 81
- ADRIAN, E D and BUYTENDIJK, F J J J Physiol, 1931, 71, 121
- BARCROFT, J and MARGARIA, R J Physiol, 1931, 72, 175 *Ibid*, 1932, 74, 156
- BATTRO, A and LABOURT, F E Rev Argent Cardiol, 1943, 10, 83
- CHRISTIE, R V and MEALINS, J J Clin Invest, 1934, 13, 323
- CHURCHILL, E D and COPE, O J Exper Med, 1929, 49, 531
- COLLIP, J B J Physiol, 1920-21, 54, 58
- COMROE, J H Physiol Rev, 1944, 24, 319
- COMROE, J H and SCHMIDT, C F Am J Physiol, 1938, 121, 75 Am J Physiol, 1943, 138, 536
- COOMBS, H. D and PIKE, F H Am J Physiol, 1918, 45, 569
- CROMER, S P and IVY, A C Proc Soc Exp Biol Med, 1931, 28, 565
- DALE, H H and EVANS, C L J Physiol, 1920, 54, 167
- DUMKE and associates, erratum see SCHMIDT and COMROE
- DOUGLAS, C G and HALDANE, J S J Physiol, 1909, 38, 401 *Ibid*, p 420
- DRINKER, C K Ann Rev Tuberc, 1948, 58, 1
- EYSTER, J A J Exp Med, 1906, 8, 565
- FAHR, G and RONZONE, E Arch Int Med, 1922, 29, 331
- FIELD, H, JR and BOCK, A V J Clin Invest, 1925, 2, 67
- FLOURENS, J P M Compt Rend Acad de Sc., Paris, 1858, 47, 803 (Year given in text incorrect, should be 1858)
- FREY, J S and GESELL, R Proc Am Soc Exp Biol 1947, 6, 106

- GESELL, R., *Physiol Rev*, 1925, 5, 551
- GRAY, J S Pulmonary Ventilation, Am Lecture Series C C Thomas, Springfield, Ill., 1949, Science, 1946, 103, 739
- GRODINS, S F See under Monographs and Reviews
- HALDANE, J B S *J Physiol*, 1921, 56, 265
- HALDANE, J S *J Physiol*, 1895, 18, 201, 430
- HALDANE, J S and POULTON, E P *J Physiol*, 1908, 37, 390
- HALDANE, J S and PRIESTLEY, J G *J Physiol*, 1905, 32, 225
- HARRISON, T R and associates Congestive heart failure The mechanism of dyspnea in exertion *Arch Int Med*, 1932, 60, 690 Studies in congestive heart failure Reflex versus chemical factors in the production of rapid breathing *J Clin Invest.*, 1932, 11, 133
- HARPOP, G A *J Am Med Assoc*, 1923, 80, 1641
- HEAD, H J *Physiol*, 1889, 10, 1 279
- HENDERSON, V E and SWEET, T A *Am J Physiol*, 1929, 91, 94
- HERING, E and BREUER, J *Sitzungsb Akad Wissensch, Wien* 1868, 57, 672 *Ibid*, 58, 909
- HEYMANS, C and BOUCKAERT, J J Sinus caroticus and respiratory reflexes I Cerebral blood flow and respiration Adrenaline apnoea. *J Physiol.*, 1930, 69, 254 Carotid sinus reflexes Influence of cerebral blood pressure and blood supply on respiratory and vasomotor centres *Ibid*, 1933, 79, 49 HEYMANS, C, BOUCKAERT, J J and DAUTREBANDE, L Sinus carotidien et réactions respiratoires au cyanure *Compt rend Soc. biol*, 1931, 106, 54. HEYMANS, J F and HEYMANS, C Activité du centre respiratoire de la tête isolée du chien *Arch. Int Pharm Therap*, 1927, 33, 273
- JACOBS, M H *Am J Physiol*, 1920, 51, 321 *ibid*, 53, 457
- LUMSDEN, T J *Physiol*, 1923, 68, 81, 111
- MARKWALD, M *Zeit f Biol*, 1887, 23, 149 *Ibid*, 1890, 26, 259
- McMICHAEL, J *Clin Sci*, 1939, 4, 19
- MEAKINS, J C *J Am Med Assoc*, 1934, 103, 1442
- NEILSON, M *Stand Arch Physiol (Suppl 10)*, 1936, 74, 87
- PARTRIDGE, G Afferent impulses in the vagus nerve. *J Cell Comp Physiol.*, 1933, 2, 367 Vagal and phrenic impulses and respiration *Can Med Assoc. J*, 1935, 33, 11
- PEMBREY, M S and ALLEN, R. W *J Physiol.*, 1905, 21, xviii
- PI SUKER, A. The regulation of the respiratory movements by chemoreceptors *Physiol Rev*, 1947, 27, 1
- PITTS, R. F The organization of the respiratory center *Physiol Rev*, 1946, 26, 609
- PITTS, R. F, MAGOUN, H. W and RANSON, S W *Am. J Physiol*, 1939, 126, 673, 689 *Ibid*, 127, 654
- PORTER, W T *J Physiol*, 1895, 17, 455
- ROSENTHAL, J *Hermann's Handb der Physiol*, 1882, 4 (2) 157
- SCHMIDT, C F and COMROE, J H *Phys Rev*, 1940, 20, 115
- SCHMIDT, C F, DUMKE, P L and DRIPPS, R D, JR *Am J Physiol*, 1939, 128, 1
- SCOTT, R W *Am J Physiol*, 1918-19, 47, 43, *ibid*, 1917, 44, 196
- SEEGERS, W H *J Biol Chem*, 1940, 136, 103
- SHARPEY SCHAFER, E J *Physiol*, 1932, 75, 130
- STELLA, G J *Physiol*, 1938, 93, 10, 263
- STEWART, G N, PIKE, F H and GUTHRIE, C C J *Exp Med*, 1908, 10, 490
- v EULER, W S and associates *Stand Arch Physiol*, 1939, 83, 132
- WATT, J G and associates *Am J Physiol*, 1943, 138, 610
- WEISS, S and ROBB, G P *J Am Med Assoc*, 1933, 100, 1841
- WINTERSTEIN, H *Pflug Arch ges Physiol*, 1911, 138, 167

Monographs and Reviews

- CORDIER, D and HEYMANS, C *Le centre respiratoire. Hermann, Paris*, 1935
- DAUTREBANDE, L Volume Jubilaire publié en l'honneur de Prof J Demoor, Thône, Liège, 1937 Quoted by Gesell, R
- GESELL, R The chemical regulation of respiration *Physiol Rev*, 1925, 5, 551 Respiration and its adjustments *Ann Rev Physiol*, 1939, 1, 185
- GRODINS, S F Analysis of Factors concerned in Regulation of Breathing in Exercise, *Physiol* 1950, 30, 220
- HALDANE, J S and PRIESTLY, J G *Respiration*, Oxford Univ Press, 1935
- HARRISON, T R. Failure of the circulation The Williams & Wilkins Co, Baltimore, 1939
- HEYMANS, C Rôle of the aortic and carotid sinus nerves in the reflex control of the respiratory center *New Eng J Med*, 1938, 219, 157
- HEYMANS, C and BOUCKAERT, J J *Le sinus carotidien et la zone homologue cardioaortique Doin, Paris*, 1933
- KROGH, A. The comparative physiology of respiratory mechanisms Univ Penn Press, Phila, 1941
- MEANS, J H. *Dyspnea Medicine*, 1924, 3, 309
- SCHMIDT, C F and COMROE, J H., JR. Functions of the carotid and aortic bodies *Physiol Rev*, 1940, 20, 115 *Ann Rev Phys*, 1941, 3, 151

CHAPTER 34

ANOXIA

- BARACH, A. L The therapeutic use of oxygen in heart disease *Ann. Int Med*, 1931, 5, 428 BARACH,

- A. L. and RICHARDS, D. W. Effects of treatment with oxygen in cardiac failure *Arch Int Med*, 1931, 48, 325
- BARCROFT, J. and associates *Phil Trans Roy Soc*, 1923, B 211, 351
- BINGER, C. A. L., BROW, G. R. and BRANCH, A. Experimental studies on rapid breathing I Tachypnea, independent of anoxemia, resulting from multiple emboli in the pulmonary arterioles and capillaries *J Clin Invest*, 1924, 1, 127 II Tachypnea, dependent upon anoxemia, resulting from multiple emboli in the larger branches of the pulmonary artery *Ibid*, p 155 BINGER, C. A. L. and MOORE, R. L. Changes in carbon dioxide tension and hydrogen ion concentration of the blood following multiple pulmonary embolism *J Exp Med*, 1927, 45, 633
- BLALOCK, A. and TAUSSIG, H. B. *J Am Med Assoc*, 1945, 128, 129
- BOYCOTT, A. E., DAMANT, G. C. C. and HALDANE, J. S. *J Hygiene*, 1908, 8, 342
- CHEN, K. K., ROSE, C. L. and CLOWES, G. H. A. *Am J Med Sci*, 1934, 188, 767
- CHRISTIE, R. V. *J Clin Invest*, 1934, 13, 295
- COMROE, J. H. and DRIPPS, R. D. *The Physiological Basis of Oxygen Therapy*, Am Lecture Series C. C. Thomas, Springfield, Ill., 1950
- DALY, I. deB. *Trans Med Chir Soc, Edin*, 1936, p 139
- DUNN, J. S. *Quart J Med*, 1920, 13, 129
- FRANKLIN, K. J. and associates *The circulation in the foetus* Blackwell Sc Pub Ltd, Oxford, 1946
- GROOM, D. and assoc. *Proc Staff Meet Mayo Clin* 1948, 23, 601
- GROSS, L. *Can Med Assoc J*, 1919, 9, 632
- HALDANE, J. S., MEAKINS, J. C. and PRIESTLEY, J. G. *J Physiol*, 1919, 52, 433
- HARROP, G. A. *J Am Med Assoc*, 1923, 80, 1641
- HENDERSON, Y. *Resuscitation J Am Med Assoc*, 1924, 83, 758 The prevention and treatment of asphyxia in the new born *Ibid*, 1928, 90, 583 The physiology of atelectasis *Ibid*, 1929, 93, 96 Observations on resuscitation from asphyxia and prevention and treatment of secondary pneumonia by inhalation of CO₂ *Brit Med J*, 1931, 2, 687 Reasons for the use of carbon dioxide with oxygen in the treatment of pneumonia *New Eng J Med*, 1932, 206, 151 HENDERSON, Y. and HENDERSON, M. C. The absorption of gas from any closed space within the body *Arch Int. Med*, 1932, 49, 88 HENDERSON, Y. and associates The treatment of pneumonia by inhalation of carbon dioxide I The relief of atelectasis *Arch Int Med*, 1930, 45, 72
- HUG, E. *Compt rend Soc biol*, 1933, 112, 511
- MEAKINS, J. and DAVIES, H. W. *J Path Bact*, 1920, 23, 451
- SAHLIN, B. *Skand Arch Physiol*, 1926, 47, 284
- SCOTT, R. W. *Arch. Int Med*, 1920, 26, 545
- SMITH, R. G. and MALCOLM, R. L. *J Pharm. Exp. Therap*, 1930, 40, 457
- STADIE, W. C. *J Exp Med*, 1922, 35, 337
- WARNER, W. P. and GRAHAM, D. *Arch. Int. Med*, 1933, 52, 888
- WENDEL, W. B. *J Am Med Assoc*, 1933, 100, 1054
- WHITEHORN, W. V. and associates *Am J Physiol*, 1946, 147, 289 *Ibid*, 1947, 148, 253 *J Aviat Med* 1947, 18, 102

Monographs and Reviews

- ABBOT, M. E. S. *Atlas of congenital cardiac disease* Amer Heart Assoc, New York, 1936
- BARCROFT, J. *The respiratory functions of the blood* Cambridge University Press, 1925
- BLALOCK, A. *Physiopathology and surgical treatment of cardiovascular defects* Harvey Lec 1945-46, 41, 90
- BURCHELL, H. B. and associates *Cardiac catheterization in the diagnosis of various cardiac malformations and diseases* Staff, Meet. Mayo Clinic, 1948, 23, 482
- CATCHPOLE, H. R. and GERSH, I. Pathogenic factors and pathological consequences of decompression sickness *Physiol Rev*, 1947, 27, 360
- COURNAND, A. and associates *Cardiac catheterization in congenital heart disease*. Commonwealth Fund, New York, 1949
- HALDANE, J. S. and PRIESTLEY, J. G. *Respiration*, Clarendon Press, Oxford, 1935
- HARVEY, N. E. *Decompression sickness* Harvey Lecture, 1943-44, 40, 41
- HILL, L. *Caisson sickness* Arnold, London, 1912
- IVY, A. C. High altitude problems in aviation *Proc Fed Am Soc Exp Biol*, 1946, 5, 319
- KOUNTZ, W. B. and ALEXANDER, H. L. *Emphysema* Medicine, 1934, 13, 251
- LUNGSGAARD, C. and VAN SLYKE, D. D. *Cyanosis* Medicine, 1923, 2, 1
- MEAKINS, J. C. and DAVIES, H. W. *Respiratory function in disease*. Oliver, Edinburgh, 1925
- SCHNEIDER, E. C. Physiological effects of altitude *Physiol Rev*, 1921, 1, 631
- VAN LIERE, E. J. *Anoxia, its effect on the body* University of Chicago Press, 1942

SECTION IV

CHAPTER 35

- URINE FORMATION, THE STRUCTURE OF THE KIDNEY, THEORIES OF RENAL FUNCTION, VOLUME AND COMPOSITION OF THE URINE
- BARCROFT, J. and STRAUB, H. *J Physiol*, 1910, 41, 145
- BAYLISS, L. E. and associates *J Physiol*, 1933, 77, 386

- BERLINER, R W and assoc. *Am J Med*, 1951, 11, 274
- BIETER, R. N and HIRSCHFELDER, A. D *Am J Physiol*, 1924, 68, 326
- BLATHERWICK, N R and LONG, M L *J Biol Chem*, 1922, 53, 103 *Ibid*, 1923, 57, 815
- BOTT, P A. and RICHARDS, A. N *J Biol Chem*, 1941, 141, 291
- BOWMAN, W *Phil Trans Roy Soc.*, 1842, 132, 57
- BRADLEY, S E See under Monographs and Reviews
- BRULL, L and EICHHOLTZ, F *Proc. Roy Soc*, 1925, B 99, 70
- BRUNN, F *Zeit ges exp Med*, 1921, 25, 170
- BRUNTON, C E and ISRAËLS, M C G BRUNTON, C E and WILSON, C *Ibid*, 1932, 74, 10 P
- CARREL, A and GUTHRIE, C C *Science*, 1906, 23, 394
- CAVETT, J W *J Biol Chem*, 1930, 87, xvi
- CHAMBERS, R and KEMPTON, R T *J Cell Comp Physiol*, 1933, 3, 131
- CUSHNY, A. R On saline diuresis *J Physiol*, 1902, 28, 431 On the secretion of acid by the kidney *Ibid*, 1904, 31, 188
- DODDS, E C Variations in alveolar carbon dioxide pressure in relation to meals *J Physiol*, 1921, 54, 342 DODDS, E C and BENNETT, T L Variations in alveolar carbon dioxide pressure in relation to meals, a further study *Ibid*, 1921, 55, 381
- DRAKIN, D L *J Biol Chem*, 1927, 75, 443 *Ibid*, p 481
- EADIE, G S and associates *Am J Physiol*, 1932, 102, 370
- EGGLETON, M G *J Physiol*, 1942, 102, 140
- ERLANGER, J and HOOKER, D R *Johns Hopkins Hosp Rep* 1904, 12, 145 See also, GESELL, R A. *Am J Physiol*, 1913, 32, 70
- GAMBLE, J L *J Biol. Chem*, 1922, 51, 295
- GOVAERTS, P *Compt. rend Soc. biol*, 1925, 93, 441
- HALDANE, J B S and associates Ammonium chloride acidosis *J Physiol.*, 1923, 57, xli. Calcium chloride acidosis *Ibid*, p 301
- HAYMAN, J M, JR. *Am J Physiol*, 1927, 79, 389, *J Clin. Invest.*, 1939, 18, 195
- HEGGIE, J F *Lancet* 1946, ii, 436, *ibid*, 1947, i, 385
- HENDERSON, L J and PALMER, W W On the intensity of urinary acidity in normal and pathological conditions *J Biol Chem*, 1913, 13, 393 On the extremes of variation of the concentration of ionized hydrogen in human urine. *Ibid*, 1913, 14, 81 On the several factors of acid excretion *Ibid*, 1914, 17, 305 On the several factors of acid excretion in nephritis *Ibid*, 1915, 21, 37
- HENDRIX, B M and SANDERS, J P *J Biol Chem*, 1925, 58, 503
- HIGGINS, H. L *Am J Physiol*, 1914, 34, 114
- HUBBARD, R. S and STEELE, T M *J Biol Chem*, 1929, 84, 199
- KALCAR, H. M *Chem Rev*, 1941, 28, 71
- KNOWLTON, F P *J Physiol*, 1911, 43, 219
- KRITZLER, R A and GRUTMAN, A B *Am J Physiol*, 1941, 134, 94
- LEATHES, J B *Brit Med Med J*, 1919, 2, 165
- LUNGGAARD, E *Biochem Ztschr*, 1933, 264, 209 221
- MAIN, R J *Am J Physiol*, 1937, 119, 7
- MARSHALL, E K, JR Agglomerular kidney of the toadfish *Bull Johns Hopkins Hosp*, 1929, 45, 95 MARSHALL, E K, JR and GRAFFLIN, A L The function of the proximal convoluted segment of the renal tubules *J Cell Comp Physiol*, 1932, 1, 161 GRAFFLIN, A L The structure of the renal tubule of the toadfish *Bull Johns Hopkins Hosp*, 1931, 48, 269
- MARSHALL, E K, JR. and VICKERS, J L *Bull Johns Hopkins Hosp*, 1923, 34, 1 Also, MARSHALL, E K *Am J Physiol*, 1931, 99
- MONKE, J V and YUTILE, C L *J Exp Med*, 1940, 72, 149
- NASH, T P, JR. and BENEDICT, S R The ammonia content of the blood, and its bearing on the mechanism of acid neutralization in the animal organism *J Biol Chem*, 1921, 48, 463 The site of ammonia formation and the rôle of vomiting in ammonia elimination *Ibid*, 1926, 69, 381 BENEDICT, S R and NASH, T P, JR. On the question of the origin of urinary ammonia *Ibid*, 1929, 82, 673
- OLIVER, J See under Monographs and Reviews
- PITTS, R F and associates *Am J Phys*, 1945, 144, 239 *J Clin Invest*, 1948, 27, 48 *Proc Fed Am Soc Exp Biol*, 1948, 7, 418, *Fed Proc* 1948, 7, 418
- PRIESTLEY, J G *J Physiol*, 1921, 55, 305
- QUINBY, W C *J Exp Med*, 1916, 23, 535
- REHBERG, P B Studies in kidney function I The rate of filtration and reabsorption in the human kidney *Biochem J*, 1926, 20, 447 II The excretion of urea and chlorine analysed according to a modified filtration reabsorption theory *Ibid*, p 461
- RICHARDS, A N and PLANT, O H. *Am J Physiol*, 1922, 59, 144
- RICHARDS, A N and SCHMIDT, C F A description of the glomerular circulation in the frog's kidney and observations concerning the action of adrenalin and various other substances upon it *Am J Physiol*, 1924, 71, 178 See also RICHARDS, A N, BARNWELL, J B and BRADLEY, R C The effects of small amounts of adrenalin upon the glomerular blood vessels of the frog's kidney perfused at constant rate *Ibid*, 1927, 79, 410
- SCHEMM, F R *Ann Int. Med*, 1942, 17, 952 *Ibid*, 1944, 21, 937
- SELLARDS, A W *Bull Johns Hopkins Hosp*, 1912, 23, 289

- SHANNON, J A *J Exp Med*, 1942, **76**, 371
- SHANNON, J A and associates *Am J Physiol*, 1941, **133**, 752
- SHEEHAN, H L *J Physiol*, 1931, **72**, 201
- SMIRK, F H and associates Studies concerning the alimentary absorption of water and tissue hydration in relation to diuresis *J Physiol*, 1932, **76**, 1 The influence of posterior pituitary hormone on the absorption and distribution of water *Ibid*, p 283 Changes in the blood composition of unanesthetized rabbits following the ingestion of water and saline *Ibid*, 1932, **76**, 81 The rate of water absorption in man and the relationship of the water load in tissues to diuresis *Ibid*, 1933, **78**, 113 The effect of water drinking on the blood composition of human subjects in relation to diuresis *Ibid*, 1933, **78**, 127 The influence of posterior pituitary hormone on the absorption and distribution of water in man *Ibid*, 1933, **78**, 147 The effect of intravenous administration of water upon the rate of urine formation *Ibid*, 1933, **78**, 451 The pituitary gland in relation to polyuria and to water diuresis *Ibid*, 1931, **31**, 172 A method for determining in animals the alimentary absorption time for water, the abdomen remaining intact *Ibid*, 1934, **81**, 167 The effect of water drinking, mineral starvation and salt administration on the total osmotic pressure of the blood in man, chiefly in relation to the problems of water absorption and water diuresis *Ibid*, 1934, **82**, 62
- SMITH, H W and associates *Am J Physiol*, 1941, **133**, 450
- STARLING, E H. *J Physiol*, 1899, **24**, 317
- STARLING, E H. and VERNEY, E B *Proc Roy Soc*, 1925, **B 97**, 321
- THEOBALD, G W *J Physiol*, 1934, **81**, 243
- THEOBALD, G W and VERNEY, E B *J Physiol*, 1935, **83**, 341
- TRUETA, J, and associates Studies of the renal circulation Thomas, Springfield, 1948
- VAN SLYKE, D D and PALMER, W W *J Biol Chem*, 1920, **41**, 567
- VAN SLYKE, D D, RHOADS, C P, HILLER, A and ALVING, A S *Am J Physiol*, 1934, **109**, 336
- VAN SLYKE, D D, STILLMAN, E and CULLEN, G E *J Biol Chem*, 1917, **30**, 401
- VERNEY, E B Absorption and excretion of water *Lancet*, 1946, **ii**, 739, 781
- VERNEY, E B and associates The secretion of pituitrin in mammals, as shown by perfusion of the isolated kidney of the dog *Proc Roy Soc*, 1926, **B 99**, 487 The absorption and excretion of water by the mammal Part I The relation between absorption of water and its excretion by the innervated and denervated kidney *Ibid*, 1933, **B 112**, 496 Part II: Factors influencing the response of the kidney to water-ingestion *Ibid*, p 521
- VERNEY, E B and WINTON, F R *J Physiol*, 1930, **69**, 153
- VIMTRUP, B *Am J Anat*, 1928, **41**, 123
- WALKER, A M and HUDSON, C L *Am J Physiol*, 1937, **118**, 130
- WALKER, A. M and OLIVER, J *Am J Physiol*, 1941, **134**, 562, 580
- WEARN, J T and RICHARDS, A N Observations on the composition of glomerular urine, with particular reference to the problem of reabsorption in the renal tubules *Am J Physiol*, 1924, **71**, 209 The concentration of chlorides in the glomerular urine of frogs *J Biol Chem*, 1925, **66**, 247 See also, RICHARDS, A. N A simple instrument for micromanipulations *Ibid*, 1930, **87**, 463 RICHARDS, A N and WALKER, A. M Quantitative studies of the glomerular elimination of phenol red and indigo carmine in frogs *Ibid*, p 479 WALKER, A. M Comparisons of total molecular concentrations of glomerular urine and blood plasma from frog and from *Necturus* *Ibid*, p 499
- WHITE, H L *Am J Physiol*, 1939, **128**, 159
- ZIMMERMAN, K W *Zeit Mikr -Anat Forsch*, 1933, **32**, 176

Monographs and Reviews

- BERGLUND, H. and MEDES, G The kidney in health and disease. Lea, Philadelphia, 1935
- BRADLEY, S E The Pathological Physiology of Uremia in Chronic Bright's Disease *Am Lecture Series* C C Thomas, Springfield, Ill, 1948
- BRUNTON, C E The acid output of the kidney and the so-called alkaline tide *Physiol Rev*, 1933, **13**, 372
- CUSHNY, A. R The secretion of urine Longmans, London, 1926
- VAN DYKE, H. B The regulation of water excretion by the neurohypophysis *Bull N Y Acad Med*, 1953, **29**, 24
- EKEHORN, G On the principles of renal function *Acta Med Scand*, Supp, **36**, 1931
- GILMAN, A AND BRAZEAU, P The role of the kidney in the regulation of acid-base metabolism *Am J Med*, 1953, **15**, 765
- GRUBER, C M The automatic innervation of the genito-urinary system *Physiol Rev*, 1933, **13**, 497
- HELLER, H. The effect of neurohypophyseal extracts on the water balance of lower vertebrates *Biol Rev*, 1945, **20**, 147
- HENDERSON, L J The excretion of acid in health and disease *Harvey Lectures*, 1914-15, **10**, 132
- MARSHALL, E K, JR The secretion of urine *Physiol Rev*, 1926, **6**, 440 The comparative physiology of the kidney in relation to theories of renal secretion. *Physiol Rev*, 1934, **14**, 133

- McCARREN, R. A. Renal function in early life. *Physiol. Rev.*, 1948, 28, 331
- MYERS, V. C. Chemical changes in the blood and their clinical significance. *Physiol. Rev.*, 1924, 4, 274.
- OLIVER, J. HARVEY. Lecture, 1944-5, 15, 765
- PARS, R. F. Renal excretion of acid. *Fed. Proc.*, 1948, 7, 418
- PHILLIPS, J. P. and VAN SLUYKE, D. D. Quantitative clinical chemistry. Vol. I. The Williams & Williams Co., Baltimore, 1931.
- RICHARDS, A. N. Methods and results of direct investigations of the function of the kidney. The Williams & Williams Co., Baltimore, 1929. Urea formation in the amphibian kidney. *Harvey Lectures*, 1934-35, 53, 93
- SHANNON, J. A. Renal tubular excretion. *Physiol. Rev.*, 1909, 19, 65
- SMITH, HOMER W. The physiology of the kidney. Oxford Univ. Press, New York, 1937. Pathology of the kidney. Porter Lecture. Univ. Kansas, Series II, 1939. Physiology of the renal circulation. *Harvey Lectures*, 1909-20 35, 166. Renal physiology between two wars. *J. Mount Sinai Hosp.*, 1945, 10, 41. Water Excretion. *N. Y. Acad. Med.*, 1947, 23, 177
- WINTON, F. R. Physiological factors in activities of Mammalian kidney. *Physiol. Rev.*, 1937, 17, 408
- CHAPTER 26
- THE PATHOLOGICAL PHYSIOLOGY OF KIDNEY DISEASE. MICROBIOLOGY
- ABEL, J. J. and associates. *J. Pharmacol.*, 1913, 5, 625
- ADDIS, T. J. *Clin. Invest.*, 1926, 2, 499
- ADDIS, T. and WATANABE, C. K. *J. Biol. Chem.* 1916, 24, 200
- ALBERT, A. and associates. *J. Lab. Clin. Med.*, 1949, 34, 326
- ALTING, A. S. and MURPHY, A. E. *J. Clin. Invest.*, 1935 15, 215
- ATKINS, J. H., SHILLMAN, E. and VAN SLUYKE, D. D. Factors governing the excretion rate of urea. *J. Biol. Chem.*, 1921, 45, 91. See also, MILLER, E., McINTOSH, J. F. and VAN SLUYKE, D. D. Studies of urea excretion. II. Relationship between urine volume and the rate of urea excretion by normal adults. *J. Clin. Invest.*, 1928, 6, 427
- III. The influence of body size on urea output. *Ibid.*, p. 467 IV. Relationship between urine volume and rate of urea excretion by patients with Bright's disease. *Ibid.*, p. 485
- BALFO, M. M. and HANAUER, J. B. S. *J. Physiol.*, 1922, 55, 259
- BARRETTTON, F. J. F. The nervous mechanism of micturition. *Quart. J. Exp. Physiol.*, 1914, 8 33 The effect of division of the hypogastric nerves on frequency of micturition. *Ibid.*, 1916, 9 261 The relation of the hind-brain to micturition. *Brain*, 1921, 44, 23 The nervous control of the urinary bladder in amphibians. *Ibid.*, 1922, 45, 126 The central nervous control of micturition. *Ibid.*, 1928, 51, 209 The component reflexes of micturition in the cat. *Ibid.*, 1931, 54, 177 The localization of the paths subserving micturition in the spinal cord of the cat. *Ibid.*, 1933, 56, 126
- BRANTON, J. and KEEL, A. S. *Brain*, 1926, 49, 302
- BECKER, E. *Ergon. d. ges. Med.*, 1935, 18, 51
- BOLLMAN, J. L. and MANN, F. C. *Proc. Soc. Exp. Biol. Med.*, 1927, 24, 925
- BOURQUIN, J. J. *Exp. Med.*, 1939, 69, 819
- BRADFORD, J. R. *J. Physiol.*, 1899, 23, 415
- CHRISTLEY, L. C. *J. Clin. Invest.*, 1937, 16, 652 *Ibid.*, 17, 119
- DEERY-BRODY, D. and POZNER, E. G. *Brain*, 1935, 58, 149
- DIEHL, H. S. and McINTOSH, C. A. In Berglund and Medes' *The kidney in health and disease*, 1935, p. 453
- EAL, D. P. and assoc. *J. Clin. Invest.*, 1944, 23, 119
- EASTMAN, A. A. Clinical types of chronic parenchymatous nephritis. *Med. Clin. Am. Ass.*, 1920, 4, 145 Further observations on the nature and treatment of chronic nephrosis. *Am. J. Med. Sci.*, 1922, 163, 167
- ERLANGER, J. and HOOKER, D. R. An experimental study of blood-pressure and pulse-pressure in man including a consideration of the relation between blood pressure and pulse-pressure and the output of albumin in a case of orthostatic albuminuria. *Johns Hopkins Hosp. Rep.*, 1904, 12, 145 See also, GRIFFIN, R. A. *Am. J. Physiol.*, 1913, 32, 70
- GOLDSCH, E. and LITTLE, J. D. *J. Clin. Invest.*, 1949, 19, 9
- GOLDING, W. and associates. Renal function in nephroses. *J. Clin. Invest.*, 1941, 20, 637
- HASTINGS, A. B. and VAN DYKE, H. B. In Berglund and Medes' *The kidney in health and disease*, 1935, p. 56.
- HENDERSON, V. E. and POFFET, M. H. *J. Pharmacol. Exp. Therap.*, 1934, 51, 97 *Ibid.*, 1935 54, 405
- HEWITT, L. F. *Biochem. J.*, 1927, 21, 1109
- HOLMES, G. *Brain*, 1933, 56, 353
- KOLFF, W. J. The artificial kidney. *J. H. Koz. N. Y. Kampen (Holland)*
- LEARNOWITZ, J. R. *Brain*, 1931, 54, 147
- LEVIN, P. M. and LANG-OLITHY, O. R. *Am. J. Physiol.*, 1937, 118, 453
- MAGLEAN, H. *Med. Res. Com.*, London, Spec. Rep. No. 43, 1949
- MAGLEAN, H. and DE WESSLOW, O. L. V. *Brit. J. Exp. Pathol.*, 1920, 1, 1
- MOORE, N. S. and VAN SLUYKE, D. D. *J. Clin. Invest.*, 1930, 8, 357
- PASTERNAK, J. H. *Arch. exp. Path. Pharmacol.*, 1910, 63, 60

- SHANNON, J A and FISHER, S *Am J Physiol*, 1938, **122**, 765
- SMITH, H W, GOLDRING, W and CHASIS, H J *Clin Invest*, 1938, **17**, 263
- TRUETA, J and associates See under Reviews and Monographs
- TUCHMAN, L R and SOBOTKA, H J *Biol Chem*, 1932, **98**, 35
- VEIL, W H *Biochem Z*, 1918, **91**, 267
- VERNEY, E B *Lancet*, 1929, **1**, 539 645 751 *Ibid*, 1930, **108**, 63
- VOLHARD, F In Berglund and Medes' *The kidney in health and disease* 1935, p 665
- WALLIS, R L M *Proc Roy Soc Med*, 1919, **13**, 96
- WINTON, F R J *Physiol*, 1934, **82**, 27 P

Monographs and Reviews

- BERGLUND, H. and MEDES, G *The kidney in health and disease* Lea, Philadelphia 1935
- BRADLEY, S E *Modern concepts of renal structure and function in Bright's disease* *N Eng J Med*, 1944, **231**, 421, *The Pathological Physiology of Uremia in Chronic Bright's Disease* C C Thomas, Springfield, Ill, 1948
- ELWYN, H *Nephritis* Macmillan, New York, 1929
- FISHBERG, A M *Hypertension and nephritis* 5th Edition, Lea and Febiger, 1944
- GRUBER, C M *The autonomic innervation of the genito-urinary system* *Physiol Rev*, 1933, **13**, 497
- HARRISON, T R and MASON, M F *The pathogenesis of the uremic syndrome* *Medicine*, 1937, **16**, 1
- HERRIN, R C *Tests of renal function* *Physiol Rev*, 1941, **21**, 529
- LEITER, L *Nephrosis* *Medicine*, 1931, **10**, 135
- MCLENNAN, F C *The neurogenic bladder* Thomas, Springfield, 1939
- OLIVER, J *Harvey Lecture*, 1944-45, **40**, 102
- TRUETA, J and associates *Studies of the renal circulation* Thomas, Springfield, 1948
- VAN SLYKE, D D and associates *Observations on the course of different types of Bright's disease and on the resultant changes in renal anatomy* *Medicine*, 1930, **9**, 257

SECTION V

CHAPTER 37

THE SALIVARY GLANDS AND THE SECRETION OF SALIVA

- ATKINSON, H F and MATHEWS, E *Brit Dental J*, 1949, **86**, 167
- DEAN, H T and associates *U S Pub Health Rep*, 1938, **53**, 1443, *ibid*, 1941, **56**, 365, *ibid*, **57**, 1155
- GIES, W J and associates *J Allied Dent Soc*, 1914, **9**, 345
- MCINTOSH, J JAMES, W W and LAZARUS-BARLOW, P *Brit J Exp Path*, 1922, **3**, 138

- PINCUS, P *Brit Dental J*, 1937, **63**, 511
- STARR, H E *J Biol Chem*, 1922, **54**, 43
- SULLIVAN, M X and DAWSON, P R *J Biol Chem*, 1921, **45**, 473
- WANG, S C *J Neurophysiol*, 1943, **6**, 195

Monographs and Reviews

- BABKIN, B P *Die aussere Sekretion der Verdauungsdrüsen* Springer Berlin, 1928
- GOTTLIEB, B *Dental caries, its etiology, pathology, clinical aspects and prophylaxis* Lea and Febiger, Phila, 1947
- LANGENSKEOLD, A *Component potentials of the sub-maxillary electrogram* Helsingfors, 1941
- MELLANBY, M *The influence of diet on the structure of the teeth* *Physiol Rev*, 1928, **8**, 545
- PAVLOV, I *The work of the digestive glands* Griffin, London, 1910
- TOVERUD, G and assoc *A Survey of the Literature of Dental Caries*, Pub 225, Nat Acad Sciences, Nat Res Counc, Wash, D C, 1952

CHAPTER 38

GASTRIC DIGESTION

- APPERLY, F L and CRABTREE, M G *J Physiol*, 1931, **73**, 331
- APPERLY, F L and CARY, M K *Am J Digest Dis Nutrit*, 1936, **3**, 466
- BABKIN, B P and KOMAROV, S A *The influence of gastric mucus on peptic digestion* *Can Med Assoc J*, 1932, **27**, 463 See also, WEBSTER, D R and KOMAROV, S A. *Mucoprotein as a normal constituent of the gastric juice* *J Biol Chem*, 1932, **96**, 133 KOMAROV, S A *Isolation of mucotinsulfuric acid from canine gastric juice* *Ibid*, 1935, **109**, 177 *The influence of mucotin-sulfuric acid on peptic digestion* *Am J Digest Dis Nutrit*, 1936, **3**, 164
- BARNES, C G *Lancet*, 1947, **2**, 549
- BROWNE, J S L and VINEBERG, A M *J Physiol*, 1932, **75**, 345
- BULL, H R and GRAY, J S *Gastroenterology*, 1945, **4**, 175
- COLLIP, J B *University Toronto Studies, Physiol Series*, No 35, 1920
- DAVENPORT, H. W *J Physiol*, 1939, **97**, 32 *Am J Physiol*, 1939, **128**, 725 *Ibid*, 1940, **129**, 505
- DRAGSTEDT, L R and ELLIS, J C 1930, **93**, 407
- FITZGERALD, M P *Proc Roy Soc*, 1910, **B 83**, 56
- FOGELSON, S J *J Am Med Assoc*, 1931, **96**, 673
- GRAY, J S *Gastroent*, 1943, **1**, 390
- HARVEY, B C H and BENSLEY, R R *Biol Bull*, 1912, **23**, 225
- HOERR, N L and BENSLEY, R R *Anat Rec*, 1936, **65**, 417
- HOLLANDER, F *The chemistry and mechanics of hydrochloric acid formation in the stomach* *Gastro-*

- enterology, 1943, 1, 401 Studies in gastric secretion, IV Variations in the chlorine content of gastric juice and their significance. *J Biol Chem*, 1932, 97, 585, Current views on the physiology of gastric secretion *Am J Med*, 51, 13, 453
- HUNT, J N *J Physiol*, 1948, 107, 365
- IVY, A C and DAWSON, A. B *Am J Physiol*, 1926, 76, 158
- IVY, A C and KIM, M S *J Am Med Assoc*, 1931, 97, 1511
- IVY, A. C and associates *Gastroent.*, 1944, 2, 138
- MALI, R Z *physiol Chem*, 1878, 1, 174.
- NORTHROP, J H. *J Gen Physiol*, 1930, 13, 739
Harvey Lectures, 1934-35, 30, 229
- Monographs and Reviews*
- BABKIN, B P The digestive work of the stomach *Physiol Rev*, 1928, 8, 365 Die aussere Sekretion der Verdauungsdrusen Springer, Berlin, 1928
- CARLSON, A. J The secretion of gastric juice in health and disease. *Physiol Rev*, 1923, 3, 1
- CONWAY, E J The Biochemistry of Gastric Secretion, Am Lecture Series C C Thomas, Springfield, Ill, 1953
- MICHAELIS, L Some problems concerning the gastric juice. Harvey Lectures, 1926-27, 22, 59
- NORTHROP, J H. and HERRIOTT, R. M Chemistry of crystalline enzymes *Ann Rev Biochem*, 1938, 7, 37
- PAVLOV, I P The work of the digestive glands Griffin, London, 1910
- WASTENEYS, H. and BORSOOK, H. The enzymatic synthesis of protein *Physiol. Rev*, 1930, 10, 110
- CHAPTER 39
- GASTRIC DIGESTION (Continued) SECRETION OF GASTRIC JUICE
- ALLEY, A. The secretory activity of the gastric mucosa in the region of the lesser curvature *Trans Roy Soc. Can*, 1933, 27, Sec V, 71 The effect upon the gastric juice secretion of various cooked preparations of haddock and of lobster *Am J Digest. Dis Nutrit*, 1934, 1, 182 ALLEY, A., MACKENZIE, D W, JR. and WEBSTER, D R. Dissociation of the functional properties of the gastric glands under the influence of fat. *Ibid*, p 333 Inhibitory effect of histamine *Ibid*, 1935, 1, 787
- ALVAREZ, W C and associates. See VANZANT, F R., ALVAREZ, W C and associates
- ANSON, M L and MIRSKY, A. E *J Gen Physiol*, 1932, 16, 59
- BABKIN, B P The value of histamine on a test of gastric secretion from a physiological point of view *Can. Med Assoc. J*, 1930, 23, 268 Does the stomach secrete gastric juice continuously? Contrib to the Med Sciences in honor of Dr Libman, Internat. Press, New York, 1932, 1, 113
- Modes of stimulation of the gastric secretion *Nature*, 1934, 134, 1005 The "chemical" phase of gastric secretion and its regulation *Am J Digest Dis Nutrit*, 1934, 1, 715
- BAIRD, M McC, CAMPBELL, J M H. and HERN, J R. B *Guy's Hosp Rep*, 1924, Series 4, 4, 23
- BARCLAY, A. E and BENTLEY, F H. *Br J Radiol.*, 1949, 22, 62, *Gastroenterol*, 1949, 12, 127
- BAXTER, S G *Am J Digest Dis Nutrit*, 1934, 1, 36 *Ibid*, p 40
- BEATTIE, J *Can Med Assoc J*, 1932, 26, 278
- BENNETT, T I and RYLE, J A *Guy's Hosp Rep*, 1921, Series 4, 1, 286
- BENNETT, T I and VENABLES, J F *Brit Med J*, 1920, 2, 662
- BIDDER, F and SCHMIDT, C Die Verdauungssafte und der Stoffwechsel Leipzig, 1952, p 45
- BLOOMFIELD, A. L *Ann Int. Med*, 1932, 6, 307
- BOLDYREFF, W *Pflüger's Arch*, 1908, 121, 13 *Quart J Exp Physiol*, 1914, 8, 1
- BOLTON, C and GOODHART, G W *Lancet*, 1922, 1, 420 *J Physiol*, 1931, 73, 115
- BONNEY, G L W and PICKERING, G W *Clin Sc.*, 1946, 6, 63
- BOWIE, D J and VINEBERG, A. M *Quart J Exp Physiol*, 1935, 25, 247
- BOYD, T E *Am J Physiol*, 1925, 71, 455 *Ibid*, p 464
- BRITTON, S W and SILVETTE, H. *Am J Physiol*, 1937, 118, 594
- BYROM, F B *J Path Bact*, 1937, 45, 1
- BUCHER, G R and IVY, A. C. *Am J Physiol.*, 1941, 132, 654
- CRIDER, J O and THOMAS, J E *Am J Physiol*, 1932, 101, 25
- DODDS, E C and associates *Lancet*, 1934, 2, 918 *Ibid*, 1935, 1, 1099
- DRAGSTEDT, L R. Supradiaphragmatic section of the vagus nerves in the treatment of gastric ulcer *Gastroenterol*, 1944, 3, 450, Removal of the vagus nerves of the stomach in gastroduodenal ulcer *Surgery*, 1945, 17, 742, Section of the vagus nerves to the stomach in the treatment of peptic ulcer *Surg, Gynec. Obstet.*, 1946, 83, 547 (Edit)
- DRAGSTEDT, L R and associates *Surg, Gynec. Obstet*, 1931, 52, 1075
- DRAGSTEDT, L R and PALMER, W L *Proc. Soc Exp Biol Med*, 1932, 29, 753
- DRAGSTEDT, L R. and VAUGHN, A M *Arch. Surg*, 1924, 8, 791
- EDKINS, J S On the chemical mechanism of gastric secretion *Proc. Roy Soc.*, 1905, B 76, 376 The chemical mechanism of gastric secretion *J Physiol*, 1906, 34, 133 EDKINS, J S and TWEEDY, M The natural channels of absorption evoking the chemical mechanism of gastric secretion *J Physiol*, 1909, 33, 263
- FARRELL, J I. *Am J Physiol*, 1928, 85, 672

- FAULEY, G B and IVY, A C *Proc. Soc. Exp Biol Med*, 1936, **34**, 152
- FITZGERALD, O *Nature*, 1946, **158**, 305, *Biochem J*, 1950, **47**, ix, FITZGERALD, O and MURPHY, P *Lancet*, ii, 1949, 1107
- GATES, F L *Proc Soc Exp Biol Med*, 1927, **24**, 936
- GAVIN, G, McHENRY, E W and WILSON, M J *J Physiol*, 1933, **79**, 234
- GILMAN, A and COWGILL, G R *Am J Physiol*, 1931, **97**, 124
- GRAY, J S and associates *Am J Physiol*, 1939, **126**, 507 *Ibid*, 1940, **129**, 589 *Proc Soc Exp Biol Med*, 1940, **43**, 225, *J Am Med Assoc*, 1951, **147**, 1529
- GREGORY, R A and IVY, A C *Quart J Exp Physiol*, 1941, **31**, 111
- GROSSMAN, M I, ROBERTSON, C R and IVY, A C *Am J Physiol*, 1948, **153**, 1
- GROSSMAN, M, WOODLEY, J R and IVY, A C *Am J Physiol*, 1944, **141**, 506
- HASTINGS, A B and assoc, quoted by Conway
- HAWK, P B, MILLER, R J and BERGEIM, O The influence of anxiety on gastric digestion *Proc Soc Exp Biol Med*, 1920, **17**, 97 HAWK, P B, MILLER, R J, BERGEIM, O and REHFUSS, M E The psychic secretion of gastric juice in normal men *Am J Physiol*, 1920, **52**, 1
- HOLLANDER, F J *Biol Chem*, 1932, **97**, 585 *Ibid*, 1934, **104**, 33, *Gastroenterology*, 1943, **1**, 401
- HURST, A F *Lancet*, 1929, **2**, 1023
- IVY, A C *Surgery*, 1941, **10**, 861
- IVY, A C and associates *Enterogastrone Am J Physiol*, 1935, **113**, 53
- IVY, A C and GRAY, J S *Enterogastrone*, Cold Spring Harbor Symposia, 1937, **5**, 405
- IVY, A C and FARRELL, J I The proof of a humoral mechanism A new procedure for the study of gastric physiology *Am J Physiol*, 1925, **74**, 639 Demonstration of stimulation of the gastric glands by mechanical distention of the stomach *Ibid*, 1925, **72**, 232
- IVY, A C and FAULEY, G B *Arch Int Med*, 1930, **46**, 52
- KELLER, A D, HARE, W K and D'AMOUR, M C *Proc Soc Exp Biol Med*, 1933, **30**, 772 *Am J Physiol*, 1934, **109**, 63
- KOCH, F C, LUCKHARDT, A. B and KEETON, R W *Am J Physiol*, 1920, **52**, 508
- KOMAROV, S A *Proc Soc Exp Biol Med*, 1938, **38**, 514, *Rev Canad Biol*, 1942, **1**, 191, 377
- LIM, R K S and associates Demonstration of the humoral agent in fat inhibition of gastric secretion *Proc Soc Exp Biol Med*, 1930, **27**, 890 On the mechanism of inhibition of gastric secretion by fat Gastric-inhibitory agent obtained from intestinal mucosa *Chin J Physiol*, 1932, **6**, 107 Observations on the mechanism of the inhibition of gastric function by fat. *Quart J Exp Physiol*, 1933, **23**, 263 Purification of enterogastrone *Chin J Physiol*, 1934, **8**, 219
- LIM, R K S, IVY, A C and MCCARTHY, J E *Quart J Exp Physiol*, 1925, **15**, 14
- LUCK, J M and SETH, T N *Bioch J*, 1924, **18**, 227, *ibid*, 1925, **19**, 357
- MCCANN, J C *Am J Physiol*, 1929, **89**, 485
- MACINTOSH, F C *Quart J Exp Physiol*, 1938, **28**, 87
- MACLEAN, H. and GRIFFITHS, W J *J Physiol*, 1928, **65**, 63 *Ibid*, 1928, **66**, 356
- MANN, F C and BOLLMAN, J L *J Am Med Assoc*, 1932, **99**, 1576
- MANN, F C and WILLIAMSON, C S *Ann Surg*, 1923, **77**, 409
- MATTHEWS, W B and DRAGSTEDT, L R *Surg, Gynec Obstet*, 1932, **55**, 265
- PALMER, W L *Arch Int Med*, 1926, **38**, 603, 694, *Ibid*, 1927, **39**, 109
- POLLAND, W S *J Clin Invest*, 1932, **11**, 449
- POPIELSKI, L *Pflüger's Arch*, 1920, **178**, 214 *Ibid*, p 237
- REHFUSS, M E, BERGEIM, O and HAWK, P B *J Biol Chem*, 1914, **19**, 345
- ROSEMANN, R *Pflüger's Arch*, 1911, **142**, 208
- SCHNEDORFF, J G and IVY, A C *J Am Med Assoc*, 1939, **112**, 898
- SHAY, H, KATZ, A B and SCHLOSS, E M *Arch Int Med*, 1932, **50**, 605
- TOBY, C G *Am J Digest Dis Nutrit*, 1937, **3**, 896
- UVNAS, B *Acta Physiol Skand*, 1942, **4**, Suppl 13 *Ibid*, **43**, 6, 97
- VANZAN, F R, ALVAREZ, W C and associates *Arch Int Med*, 1932, **49**, 345
- VINEBERG, A M *Am J Physiol*, 1931, **96**, 363
- WILHELMJ, C M and associates *Am J Physiol*, 1936, **116**, 685
- WILHELMJ, C M and associates *Am J Physiol*, 1937, **118**, 766 *Ibid*, **120**, 619
- WILSON, M J *Arch Int Med*, 1928, **41**, 633

Monographs and Reviews

- BENNETT, T I, *Diseases of the upper alimentary tract* Heinemann, London, 1925
- BLOOMFIELD, A L and POLLAND, W S *Gastric anacidity, its relation to disease* Macmillan, New York, 1933
- CARLSON, A J *The control of hunger in health and disease (psychic secretion in man)* Chicago University Press, 1916
- CUSHING, H *Peptic ulcer and the interbrain* (Fourth Balfour Lecture, University of Toronto) In, *The pituitary body, hypothalamus and parasympathetic nervous system* Thomas, Springfield, 1932
- GREENGARD, H *Hormones of the Gastrointestinal Tract*, in *The Hormones*, G PINCUS and K V THIMANN, Ed Academic Press, N Y, 1948

- GROSZKA, M. I. *Gastromtestinal hormones* Physiol. Rev, 1950, 30, 33
- HOLLANDER, F and PENNER. *History of gastric analysis*. Am. J. Dig. Diseases, 1938, 5, 739, 786 Ibid, 1939, 6, 22, Am. J. Med, 1952, 13, 453
- IVY, A. C. *The rôle of hormones in digestion*. Physiol. Rev, 1930, 10, 282 *The gastromtestinal hormones*. Trans. and Studies. Coll. Physicians, Philadelphia, 1944, 12, 101
- IVY, A. C. and GRAY, J. S. *Enterogastrone*. Cold Spring Harbor Symposia Quant. Biol., 1937, 5, 405
- KINSELLA, V. J. *The mechanism of abdominal pain*. Australasian Med. Pub. Co., Sydney, 1948
- RYLE, J. A. *Gastric function in health and disease*. Milford, London, 1926
- Series of papers by various authors on production of peptic ulcer. Arch. Surg, 1942, 44, 399-530
- WOLF, S. and WOLFF, H. G. *Human gastric function*. Oxford Univ. Press, 1943
- CHAPTER 40
- DIGESTION IN THE INTESTINE
- AGREN, G. and associates. Skand. Arch. Physiol., 1934, 69, 1 Acta Med. Skand., 1936, 90, 1 Ibid, 1937, 100
- ANTREP, G. VON. J. Physiol., 1914, 49, 1
- ARCHIBALD, E. Surg., Gynec. Obstet., 1919, 28, 529 *Acute pancreatitis*. Nelson's Looseleaf Medicine, 1920, 5, 536
- BAYLISS, W. M. and STARLING, E. H. J. Physiol., 1902, 28, 325
- BERMAN, A. L. and associates. Quart. J. Studies Alcohol, 1941, 1, 645
- BLANKENHORN, M. A. and HOOVER, C. F. Arch. Int. Med., 1916, 18, 289
- BLOCK, K. and associates. J. Biol. Chem., 1943, 149, 511 Ibid, 1945, 157, 601
- CARE, J. L. and FOOTE, F. S. Arch. Surg., 1934, 29, 277
- COLLISON, G. A. and FOWWEATHER, F. S. Brit. Med. J., 1926, 1, 1081
- CRIDER and THOMAS, see THOMAS and CRIDER.
- DAM, H. and GLAVIND, J. Lancet, 1938, 1, 720
- DOLKART, R. E., JONES, K. K. and BROWN, C. F. G. Arch. Int. Med., 1938, 62, 618
- DORCHESTER, J. E. C. and HAST, R. E., J. Physiol., 1952, 118, 182, Ibid., 118, 188
- DRAGSTEDT, L. R., HAYMOND, H. E. and ELLIS, J. C. Arch. Surg., 1934, 28, 232
- DRILL, V. A. and IVY, A. C. J. Clin. Invest., 1944, 23, 209
- DRURY, D. R., McMASTER, P. D. and ROUS, P. J. Exp. Med., 1924, 39, 403
- ELMAN, R. and McCaUGHAN, J. M. J. Exp. Med., 1927, 45, 561
- ELMAN, R. and McMASTER, P. D. *Studies on urobilin physiology and pathology I. The quantitative determination of urobilin*. J. Exp. Med., 1925, 41, 503 *II. Derivation of urobilin*. Ibid, p. 513 *III. Absorption of pigments of biliary derivation from the intestine*. Ibid., p. 719 *IV. Urobilin and the damaged liver*. Ibid., 1925, 42, 99 *V. The relation between urobilin and conditions involving increased red cell destruction*. Ibid, p. 619 *VI. The relation of biliary infections to the genesis and excretion of urobilin*. Ibid, 1926, 43, 73
- FLOREY, H. W. Brit. J. Exp. Path., 1932, 13, 349
- FLOREY, H. W. and HAEDING, H. E. Proc. Roy. Soc., 1935, B 117, 68
- GAMBLE, J. L. and McIVER, M. A. J. Exp. Med., 1928, 48, 859
- GARDNER, J. C. and GAINSBOROUGH, H. Quart. J. Med., 1930, 23, 465
- GRAY, S. J. Arch. Int. Med., 1940, 65, 524
- GREAVES, J. D. and SCHMIDT, C. L. A. Proc. Soc. Exp. Biol. Med., 1932, 29, 373 Ibid, 1937, 37, 40 Am. J. Physiol., 1935, 111, 492
- GREENGARD, H. and associates. Science, 1944, 99, 350
- GROSSMAN, M. I., GREENGARD, H. and IVY, A. C. Am. J. Physiol., 1943, 138, 676, Ibid, 1944, 141, 38
- HANGER, F. M. J. Clin. Invest., 1939, 18, 261
- HARPER, A. A. and RAPER, H. S. J. Physiol., 1943, 102, 115
- HARPER, A. A. and VASS, C. C. V. J. Physiol., 1941, 99, 415
- HUNTER, G. Brit. J. Exp. Path., 1930, 11, 407 Ibid, p. 415
- IVY, A. C. *Certain aspects of the applied physiology of the external secretion of the pancreas*. Am. J. Digest. Dis. Nutrit., 1936, 3, 677
- IVY, A. C. and CRITTENDEN, P. J. *The nervous control of pancreatic secretion in the dog*. Am. J. Physiol., 1937, 119, 724
- IVY, A. C., FARRELL, J. I. and LUETH, H. C. Am. J. Physiol., 1927, 82, 27
- JEMERIN, E. E. and HOLLANDER, F. Proc. Soc. Exp. Biol. Med., 1938, 38, 139
- KUNITZ, M. J. Gen. Physiol., 1938, 22, 477
- LEMBERG, R. Ann. Rev. Biochem., 1936, 5, 479, Biochem. J., 1935, 29, 1322 Ibid, 1938, 32, 149, 171
- MACLAGAN, N. F. Br. J. Exp. Path., 1944, 25, 234
- McMASTER, P. D. and ROUS, P. *The biliary obstruction required to produce jaundice*. J. Exp. Med., 1921, 33, 731 *ROUS, P. and McMASTER, P. D. Physiological causes for the varied character of stasis bile*. J. Exp. Med., 1921, 34, 75
- McNEE, J. W. J. Path. Bact., 1913, 18, 325
- MANA, F. C. and associates. *The formation of bile pigment after total removal of the liver*. Am. J. Physiol., 1924, 69, 393 *The site of formation of bilirubin*. Ibid, 1925, 74, 497 *The formation of bile pigment from hemoglobin*. Ibid, 1926, 76, 306 *Spectrophotometric determination of bili-*

- rubin Ibid , 1926, 76, 577 The liver as a site of bilirubin formation Ibid , 1926, 77, 219 An experimental study of reduced hepatic function Ibid , p 179 An evaluation of the relative amounts of bilirubin formed in the liver, spleen and bone marrow Ibid , 1926, 78, 384 A method for making a satisfactory fistula at any level of the gastrointestinal tract Ann Surg , 1931, 93, 794
- MARKOWITZ, J and HOUGH, H B Am J Physiol , 1926, 75, 571
- MEAKINS, J J Clin Invest , 1927, 4, 135
- MELLANBY, J The secretion of pancreatic juice J Physiol , 1926, 61, 419 Secretion and the portal circulation Ibid , p 489 Bile salts and secretin as cholagogues Ibid , 1928, 64, 331 Isolation of secretin, its chemical title and physiological properties Ibid , 1928, 66, 1
- MELLANBY, J and HUGGETT, A St G J Physiol , 1926, 61, 122
- MINKOWSKI, I and NAUNYN, B Arch exp Path Pharm , 1886, 21, 1
- NASSET, E S Am J Physiol , 1938, 121, 481
- OKDA, S J Physiol , 1915, 50, 114
- QUICK, A J, STANLEY-BROWN, M and BANCROFT, F W Am J Med Sci , 1935, 190, 501
- RABINOWITCH, I M Arch Int Med , 1930, 46, 1014
- RAMSAY, A J, THOMAS, J E and CRIDER, J O Anat Rec , 1943, 86, 87
- RICH, A R Bull Johns Hopkins Hosp , 1924, 35, 415, Ibid , 1930, 47, 338
- RICH, A R and DUFF, G L Bull Johns Hopkins Hosp , 1936, 58, 137
- RODERICK, L M J Am Vet Med Assn , 1929, 74, 314
- ROWNTREE, L G, GREENE, C H and ALDRICH, M J Clin Invest , 1927, 4, 545
- SCHEPOWALNIKOW, N P Dissertation St Petersburg , 1899 Quoted by Babkin
- SHERLOCK, S P V Lancet, 1945, II, 397 Ibid , J Path and Bact , 46, 58, 523
- SNYDER, C D Am J Physiol , 1937, 118, 345
- TANTURI, C H and IVY, A C Am J Physiol , 1938, 121, 270
- TAYLOR, N B, WELD, C B and SYKES, J F Brit J Exp Path , 1935, 16, 302
- THOMAS, J E Proc Federation Am Soc Exp Biol , 1942, 1, 261
- THOMAS, J E and CRIDER, J O Am J Physiol , 1941, 134, 656, Am J Med Sci , 1942, 203, 155, Am J Physiol , 1944, 141, 730
- VAN DEN BERGH, A. A H. Der Gallenfarbstoff im Blut Leiden, 1918 See also Discussion on jaundice Brit Med J , 1924, 2, 498
- VANZANT, F R, ALVAREZ, W C and associates Arch Int Med , 1932, 49, 345
- VERZÁR, F Am J Physiol , 1929, 90, 546
- VIRTUE, R W and DOSTER-VIRTUE, M E J Biol Chem , 1939, 127, 431
- WALTON, SIR JAS Br Med J , 1945, II, 593
- WATSON, C J Blood, 1946, 1, 99
- WHIPPLE, G H and HOOPER, C W J Exp Med , 1913, 17, 612
- WRIGHT, R D and associates Quart J Exp Physiol , 1940, 30, 73

Monographs and Reviews

- ASCHOFF, L The origin of gall-stones Hoeber, New York, 1924
- BABKIN, B P Secretory mechanism of the digestive glands Hoeber, New York, 1944
- BARRON, E S G Bilirubinemia Medicine, 1931, 10, 77
- FLOREY, H W, WRIGHT, R D and JENNINGS, M A. The secretions of the intestine Physiol Rev , 1941, 21, 36
- GRAHAM, E A, COLE, W H, COPER, G H, and MOORE, S Diseases of the gall-bladder and bile ducts Lea, Philadelphia, 1928
- GREENGARD, H Hormones of the gastrointestinal tract, in The Hormones, G PINCUS and K V THIMANN, Ed Academic Press, N Y , 1948
- GROSSMAN, M I Gastrointestinal hormones Physiol Rev , 1950, 30, 33
- IVY, A C The rôle of hormones in digestion Physiol Rev , 1930, 10, 282
- MACLEAN, H. Modern views on digestion and gastric disease Constable, London, 1928
- MCNEE, J W Jaundice, a review of recent work Quart J Med , 1923, 16, 390
- MCNEE, J W Discussion on jaundice Brit Med J , 1924, 2, 496
- MANN, F C The effects of complete and partial removal of the liver Medicine, 1927, 6, 419
- PATEY, D H Modern views on the mechanism of gall-stone formation Brit Med J , 1933, 1, 866
- RICH, A. R. The formation of bile pigment Physiol Rev , 1925, 5, 182
- ROLLESTON, H. D and MCNEE, J W Diseases of the liver, gall bladder and bile ducts Macmillan, London, 1929
- SCHMIDT, C L A The extra-hepatic functions of bile Physiol Rev , 1927, 7, 129
- SOFFER, L J Present day status of liver function tests Medicine, 1935, 14, 185
- SOBOTKA, H The chemistry of the bile acids and related substances Chem Rev , 1934, 15, 311 The physiological chemistry of bile The Williams & Wilkins Co , Baltimore, 1937
- STILL, E U Secretin Physiol Rev , 1931, 11, 328
- THOMAS, J E The External Secretion of the Pancreas Am Lecture Series C C Thomas, Springfield, Ill , 1950
- VERZAR, F and MCDUGALL, E J Absorption from the intestine Longmans, Lond , 1936
- WAKIM, K G The physiology of the liver Am J Med , 1954, 16, 256

- WALDSCHMIDT-LEITZ, E Enzyme actions and properties Wiley, New York, 1929
- WATSON, C J New Eng J Med, 1942, 227, 99, *ibid* 1947, 237, 225, 261, Blood, 1946, 1, 225, 261
- WHIPPLE, G H The origin and significance of the constituents of the bile Physiol Rev, 1922, 2, 440

CHAPTER 41

THE GALL BLADDER AND BILE DUCTS

- BOYDEN, E A and associates Anat Rev, 1923, 24, 388 Anat Rec, 1928, 40, 147 Surgery, 1937, 1, 25 *Ibid*, 1941, 10, 567 *Ibid*, 1941, 9, 443 *Ibid*, 1934, 69, 427 *Ibid*, 1940, 76, 485
- DRURY, D R and ROUS, P J Exp Med, 1925, 41, 611
- ELMAN, R. and GRAHAM, E A. Arch Surg, 1932, 24, 14
- ELMAN, R. and TAUSSIG, J B J Exp Med, 1931, 54, 775
- HIGGINS, G M and MANN, F C Am J Physiol, 1926, 78, 339
- IVY, A. C. and FISHBACK Bull N Y Acad Med, 1934, 10, 643
- IVY, A. C. and OLDBERG, E Am J Physiol, 1928, 86, 599
- JOHNSON, F E and BOYDEN, E A. Surg, Gyn Obstet, 1943, 78, 395
- KEITH, A Lancet, 1903, 1, 631
- LICHTENSTEIN, M E and IVY, A. C Surgery, 1937, 1, 38
- LUETH, H C Am J Physiol, 1931, 99, 237
- MANN, F C and BOLLMAN, J L J Lab Clin Med, 1925, 10, 540
- ROUS, P and MCMASTER, P D J Exp Med, 1921, 34, 47
- WHITAKER, L R The mechanism of the gall bladder Am J Physiol, 1926, 78, 411 KRAUSE, W F and WHITAKER, L R. Effects of different food substances upon emptying of the gall bladder *Ibid*, 1928, 87, 172

Monographs and Reviews

- GRAHAM, E A. The clinical application of some recent knowledge of the biliary tract Harvey Lectures, 1933-34, 29, 176
- IVY, A. C The physiology of the gall bladder Physiol Rev, 1934, 14, 1
- MANN, F C The functions of the gall bladder Physiol Rev, 1924, 4, 251

CHAPTER 42

THE MOVEMENTS OF THE ALIMENTARY CANAL

- ALBRECHT, F K Med Clin N Am, 1943, 27, 1052
- ALVAREZ, W C The length of the interval between the eating of a food and the appearance of the

- distress caused thereby Proc Staff Meetings, Mayo Clinic, 1935, 10, 103 Ways in which emotion can affect the digestive tract J Am Med Assoc, 1929, 92, 1231
- APPERLY, F L Brit J Exp Path, 1926, 7, 011
- BAIRD, M Mc, CAMPBELL, J M H and HERN, J R B Guy's Hosp Rep, 1924, Series 4, 4, 23
- BARCLAY, A E Lancet, 1934, 1, 11 Proc. Staff Meetings, Mayo Clinic, 1930, 5, 251
- BARD, P and assoc Fed Proc, 1947, 6, 72
- BEATTIE, J and SHEEHAN, D J Physiol, 1934, 81, 218
- CANNON, W B Movements of the stomach studied by means of Röntgen rays Am J Physiol, 1898, 1, 359 Emptying of the human stomach *Ibid*, 1904, 10, 111
- CANNON, W B and WASHBURN, A L Am J Physiol, 1912, 2, 441
- CARLSON, A J, LUCKHARDT, A B and PHILLIPS, H T Am J Physiol, 1910, 50, 57
- COLE, L G The value of serial radiography in gastro intestinal diagnosis J Am Med Assoc, 1912, 59, 1947 Motor phenomena of the stomach, pylorus and cap observed roentgenographically Am J Physiol, 1917, 42, 618 (Proc) The living stomach and its motor phenomenon Acta Radiologica, Stockholm, 1928, 9, 533
- DRAGSTEDT, L R Surg, Gynec Obstet, 1931, 52, 105
- GAY, L N and CARLINER, P E Science, 1949, 109, 359
- GORDON, S and SINGLETON, A C Surgery, 1939, 6, 697
- GRIMSON, K S and associates Surg, 1946, 20, 94
- HARDING, V J Lancet 1921, 2, 327 HARDING, V J and WATSON, B P Lancet, 1922, 2, 649
- HATCHER, R A and WEISS, S J Pharm Exp Therap, 1923, 22, 139
- HWANG, K and associates Am J Physiol, 1947, 149, 329
- JONES, C M and RICHARDSON, W J Clin Invest., 1926, 2, 610 (Proc)
- KAST, L Berl Klin Woch, 1906, 43, 947
- KELLING Arch Chr, 1901, 64, 393 (Quoted by Alvarez)
- KLEINE, E Arch Surg, 1926, 12, 571
- KNIGHT, G C Brit J Surg, 1934, 22, 155 *Ibid*, p 864
- MCCANN, J C Am J Physiol, 1929, 89, 497
- MCCLURE, C W, REYNOLDS, L and SWARTZ, C O Arch Int Med, 1920, 26, 410
- MCCRAE, E D'A J Anat, 1924, 59, 18
- MCCREA, E D and McSWINEY, B A The effect of stimulation of the vagi on the pyloric region of the stomach J Physiol, 1926, 61, 28 The effect on the stomach of stimulation of the peripheral end of the splanchnic nerve Quart J Exp Physiol, 1928, 18, 301 McCREA, E D, McSWINEY, B A and STOPFORD, J S B Stimula-

- tion of the peripheral end of the vagus *Quart Exp Physiol*, 1925, 15, 201
- MACLEOD, J J R, MAGEE, H E and PURVES, C B *J Physiol*, 1930, 70, 404
- McSWINEY, B A and SPURRELL, W R *J Physiol*, 1933, 79, 437
- PAYNE, W W and POULTON, E P *Quart J Med*, 1923, 17, 53
- POULTON, E P *Lancet*, 1928, 2, 1223, 1277
- RAKE, G W *Guy's Hosp Rep*, 1927, Series 4, 7, 141
- SHEEHAN, D J *Physiol*, 1934, 83, 177
- STRICKLAND, B A, JR and HAHN, G L *Science*, 1949, 109, 359
- THOMAS, J E, CRIDER, J O and MORGAN, C J *Am J Physiol*, 1934, 108, 683
- VAN LIERE, E J and NORTHRUP, D W *Gastroenterol*, 1943, 1, 279
- VAN LIERE, E J and SLEETH, C K *Am J Dig Dis*, 1940, 7, 118
- WANG, S C and BORISON, H L *Abstracts, XVIII, Internat Congress Physiol*, Copenhagen, 1950, *Am J Physiol*, 1951, 164, 520, *ibid*, 1951, 166, 712, *Arch Neurol Psychiat*, 63, 928, *Gastroenterol*, 1952, 22, 1
- WELCH, B J *Am Med Assoc*, 1945, 129, 204
- WHEELON, H and THOMAS, J E Observation on the motility of the antrum and the relation of rhythmic activity of the pyloric sphincter to that of the antrum *J Lab Clin Med*, 1920, 6, 124 Rhythmicity of the pyloric sphincter *Am J Physiol*, 1921, 54, 460 Observations on the motility of the duodenum and the relation of duodenal activity to that of the pars pylorica *Ibid*, 1922, 59, 72
- THOMAS, J E A further study of the nervous control of the pyloric sphincter *Am J Physiol*, 1929, 88, 498
- WILSON, M J and IRVING, L *Can Med Assoc J*, 1931, 25, 685
- ment, experimental observations MacMillan, New York, 1938
- McSWINEY, B A. Innervation of the stomach *Physiol Rev*, 1931, 11, 478
- MURLIN, J R The emptying mechanism of the stomach *J Nutrit*, 1930, 2, 311
- NOBLE, R L Treatment of experimental motion sickness in humans *Can J Research*, 1946, Feb, p 12 Motion sickness with special reference to air sickness *Practitioner*, 1948, 160, 453
- TODD, W T Behavior patterns of the alimentary tract The Williams & Wilkins Co, Baltimore, 1933
- TYLER, D B and BARD, P *Physiol Rev*, 1949, 29, 281

CHAPTER 43

MOVEMENTS OF THE ALIMENTARY CANAL (*Continued*)
 ABSORPTION FROM THE INTESTINAL TRACT, THE
 FORMATION OF FECES, CONSTIPATION,
 INTESTINAL OBSTRUCTION

- ALVAREZ, W C The metabolic gradient underlying peristalsis *J Am Med Assoc*, 1919, 73, 1438 Also, series of papers by Alvarez, and by Alvarez and Starkweather, in *Am J Physiol*, from 1915, vol 37 to 1919, vol 50
- ALVAREZ, W C and MAHONEY, L J The myogenic nature of the rhythmic contractions of the intestine *Am J Physiol*, 1922, 59, 421 See also, GASSER, H S *J Pharm Exp Therap*, 1926, 27, 395, and THOMAS, J E and KUNTZ, A *Am J Physiol*, 1926, 76, 606
- BARCLAY, A E *Lancet*, 1934, 1, 11
- BAYLISS, W M, and STARLING, E H *J Physiol*, 1899, 24, 99 *Ibid*, 1901, 26, 125
- BRODIE, T G, CULLIS, W C and HALLIBURTON, W D *J Physiol*, 1910, 40, 173
- CAMERON, J A M *Arch Dis Child*, 1928, 3, 210
- CANNON, W B *Am J Physiol*, 1902, 6, 251
- COPE, C and associates *J Clin Invest*, 1943, 22, 111
- CORI, C F *J Biol Chem*, 1925, 66, 691
- DRAGSTEDT, L R and ELLIS, J C *Am J Physiol*, 1930, 93, 407
- FALCONER and associates *Proc Staff Meet. Mayo Clinic*, 1939, 14, 22
- FINE, J and associates *Ann Surg*, 1940, 112, 546
- GAMBLE, J L and ROSS, S G *J Clin Invest*, 1924, 1, 403
- GARRY, R C *J Physiol*, 1933, 77, 422
- GREENWOOD, W F, HAIST, R E and TAYLOR, N B *Surgery*, 1940, 7, 280
- GUNN, J A and UNDERHILL, S W F *Quart J Exp Physiol*, 1914, 8, 275
- HADEN, R L and ORR, T G *J Exp Med*, 1923, 38, 55 *Arch Surg*, 1925, 11, 859
- HANKE, M T and KOESSLER, K K *J Biol Chem*, 1924, 59, 889
- HARTWELL, J A and HOUGUET, J P *J Am Med Assoc*, 1912, 59, 82
- Monographs and Reviews*
- ALVAREZ, W C The mechanics of the digestive tract Hoeber, N Y, 1928
- ALVAREZ, W C *Nervous indigestion* Hoeber, New York, 1931
- BARCLAY, A E The digestive tract Cambridge University Press, 1933
- BORISON, H L and WANG, S C The physiology and pharmacology of vomiting *Pharmacol Rev*, 1953, 5, 193
- CANNON, W B The mechanical factors of digestion Arnold, London, 1911
- CARLSON, A J The control of hunger in health and disease University of Chicago Press, 1917
- EVANS, C L The physiology of plain muscle *Physiol Rev*, 1926, 6, 358
- HATCHER, R A The mechanism of vomiting *Physiol Rev*, 1924, 4, 479
- JONES, C M Digestive tract pain diagnosis and treat-

- HASTINGS, A. B., MURRAY, C. D. and MURRAY, H. A., *J. Biol. Chem.*, 1921, 46, 223
- HENDERSON, V. E. *Am. J. Physiol.*, 1928, 86, 82
- HERZL, R. C. and MEER, W. J. *Arch. Int. Med.*, 1933, 51, 152
- HEVITT, J. A. *Biochem. J.*, 1924, 18, 161
- IVY, A. C. and associates. *Am. J. Physiol.*, 1938, 121, 137 *Am. J. Dig. Dis.*, 1940, 7, 500
- HÖBER, R. *Pflüger's Arch.*, 1904, 101, 607
- KARL, L. *Gastric Absorption. Physiol. Rev.*, 1948, 28, 433
- KEITH, N. M. and BINGER, M. W. *J. Am. Med. Assoc.*, 1935, 105, 1584
- LEARMONTH, J. R. and MARKOWITZ, J. *Am. J. Physiol.*, 1930, 94, 501 *Ibid.*, 1929, 89, 686
- LEARMONTH, J. R. and RANER, F. W. *Ann. Surg.*, 1930, 92, 710
- MACLEOD, J. J. R., MAGEE, H. E. and PERVES, C. B. *J. Physiol.*, 1930, 70, 404. *ACHENACHIE, D. W., MACLEOD, J. J. R. and MAGEE, H. E. Ibid.*, 1930, 69, 185
- MALL, J. *Johns Hopkins Hosp. Rep.*, 1896, 93, 256
- MARKOWITZ, J. and CAMPBELL, W. R. *Am. J. Physiol.*, 1927, 81, 101
- MEAKINS, J. and HARRINGTON, C. R. *J. Pharm. Exp. Therap.*, 1921, 18, 455 *Ibid.*, 1922, 20, 45
- MILLER, T. G. and ABBOTT, W. O. *Am. J. Med. Sci.*, 1934, 187, 595
- QUIGLEY, J. P., HIGHTSTONE, W. H. and IVY, A. C. *Am. J. Physiol.*, 1934, 108, 151 *REID, P. E., IVY, A. C. and QUIGLEY, J. P. Ibid.*, 1934, 109, 483
- ROBERTSON, E. C. and DOYLE, M. E. *J. Nutrition*, 1935, 9, 553
- SCOTT, W. J. M. and MORTON, J. J. *J. Clin. Invest.*, 1931, 9, 247
- STICKNEY, J. C. and associates. *Am. J. Physiol.*, 1947, 150, 466
- SWENSON, O. and assoc. *Surgery*, 1948, 24, 212, *Am. J. Physiol.*, 1948, 155, 439, *New Eng. J. Med.*, 1949, 241, 551
- TAYLOR, N. B., WEID, C. B. and HARRISON, K. *Can. Med. Assoc. J.*, 1933, 29, 227
- VAN LIERE, E. J. and associates. *Am. J. Physiol.*, 1940, 129, 618 *Ibid.*, 1941, 134, 288
- VAN STICKNEY and associates, see STICKNEY and associates.
- VISSCHER, M. B. and associates. *Am. J. Physiol.*, 1944, 142, 550 *Ibid.*, 1945, 144, 457
- WA. GENSTEIN, O. H. and PAINE, J. R. *J. Am. Med. Assoc.*, 1933, 101, 1532
- WEISS, S., ROBB, G. B. and ELLIS, L. B. *Arch. Int. Med.*, 1932, 49, 360
- WELLS, H. S. *Am. J. Physiol.*, 1931, 99, 209, *ibid.*, 1940, 150, 410
- WHIFFLE, G. H. and associates. *Bull. Johns Hopkins Hosp.*, 1912, 23, 159 *J. Exp. Med.*, 1913, 17, 285, 307 *Ibid.*, 1914, 19, 144. *Ibid.*, 1916, 23, 123
- Contributions to medical and biological research, dedicated to Sir William Osler Hoerber, New York, 1919
- ZUCKER, T. F. and MATZNER, M. J. *Proc. Soc. Exp. Biol. Med.*, 1923, 21, 186 See also, JEFFCOTT, H. and BACHRACH, L. *Biochem. J.*, 1926, 20, 1351, and TISDALL, F. F. and PRICE, H. W. *Bull. Johns Hopkins Hosp.*, 1927, 41, 432
- ZWEMER and associates. *Ann. Surg.*, 1938, 107, 161

Morographs and Reviews

- ALVAREZ, W. C. *Intestinal auto-intoxication. Physiol. Rev.*, 1924, 4, 352
- AMBROSE, A. M. and SHEEVIN, C. P. *Detoxication mechanisms. Ann. Rev. Biochem.*, 1933, 2, 377
- BARCLAY, A. E. *The digestive tract. Cambridge University Press*, 1933
- BAUGARTNER, C. J. *Survey of intestinal obstruction. Arch. Surg.*, 1947, 55, 607
- CODE, C. F. and assoc. *Motility of the alimentary canal in man. Am. J. Med.*, 1952, 13, 75
- GARRY, R. C. *The movements of the large intestine. Physiol. Rev.*, 1934, 14, 103
- GASK, G. E. and ROSS, J. P. *The surgery of the sympathetic nervous system. Wood, Baltimore*, 1934.
- HANDLER, P. and PERLZWEIG, W. A. *Detoxication mechanisms. Ann. Rev. Biochem.*, 1945, 16, 617
- HURST, A. F. *Constipation and allied disorders. Frowde, London*, 1919
- MAGEE, H. E. *The rôle of the small intestine in nutrition. Physiol. Rev.*, 1930, 10, 473
- TODD, T. W. *Behaviour patterns in alimentary canal. The Williams & Wilkins Co., Baltimore*, 1930
- VAN LIERE, E. J. *The effect of anoxia on the alimentary tract. Physiol. Rev.*, 1941, 21, 307
- VERZAR, F. and McDONOUGH, E. J. *Absorption from the intestine. Longmans, Lond.*, 1936

CHAPTER 44

VISCERAL SENSATIONS

- ADOLPH, C. F. *Am. J. Physiol.*, 1947, 151, 110
- BARCLAY, A. E. *Lancet*, 1922, 2, 261
- BENTLEY, F. H. and SMITHWICK, R. H. *Lancet*, 1940, II, 359
- BOLTON, C. *Lancet*, 1928, 1, 1159, 1211, 1263
- CANNON, W. B. *Proc. Roy. Soc.*, 1917, B 90, 283
- CARLSON, A. J. and HOELZEL, F. *Science*, 1952, 116, 526
- GILMAN, A. *Am. J. Physiol.*, 1937, 120, 323
- GREGersen, M. L. and CANNON, W. B. *Am. J. Physiol.*, 1932, 102, 336
- HEAD, H. *Brain*, 1893, 16, 1 *Ibid.*, 1894, 17, 339 *Ibid.*, 1896, 19, 153
- JONES, C. M. *Digestive Tract Pain. Macmillan, New York*, 1938
- KELLGREN, J. H. *Visceral pain. Clin. Sc.*, 1938, 3, 175
- LEWIS, T. and KELLGREN, J. H. *Clin. Sc.*, 1939, 4, 47

- MONTGOMERY, M F *Am J Physiol.*, 1931, 96, 221
 Ibid , 1931, 98, 35 See also GREGERSEN, M L
 and CANNON, W B Ibid , 1932, 102, 336
 POLLOCK, L J and DAVIS, L *Arch Neurol and Psy-*
chiat , 1935, 34, 1041
 POULTON, E P *Lancet*, 1928, 2, 1223, 1277 PAYNE,
 W W and POULTON, E P *J Physiol*, 1927, 63,
 217
 ROSS, J *Brain*, 1888, 10, 333
 ROWNTREE, S G and associates *Arch Int Med* ,
 1922, 29, 306
 RYLE, J A *Lancet*, 1926, 1, 895
 SHEEHAN, D J *Anat* , 1933, 67, 233
 WOOLLARD, H H and CARMICHAEL, E A *Brain*,
 1935, 56, 293

Monographs and Reviews

- ADOLPH, E F and associates *Man in the Desert*.
 Interscience Pub , New York, 1947
 CANNON, W B *Hunger and thirst Murchison's hand-*
book general Exp psychol Clark Univ Press,
 Worcester, Mass , 1934
 CARLSON, A J *The control of hunger in health and*
disease Chicago University Press, 1916
 HURST, A F *The sensibility of the alimentary canal*
 Frowde, London, 1911
 KINSELLA, V J *The mechanism of abdominal pain*
 Australasian Med Pub Co , Sydney, 1948
 LENNANDER, K G *Observations on the sensibility*
of the abdominal cavity Bale, London, 1903
 (Quoted by Morley)
 LEWIS, SIR T *Pain* Macmillan, N Y , 1942
 MACKENZIE, J *Symptoms and their interpretations*
 Shaw, London, 1920
 MAYEP, J *Genetic, traumatic and environmental*
factors in etiology of obesity *Physiol Rev* , 1953,
 33, 472
 MORLEY, J *Abdominal pain* Livingstone, Edinburgh,
 1931
 RYLE, J A *Gastric functions in health and disease*
 Milford, London, 1926

SECTION VI

CHAPTER 45

GENERAL METABOLISM

- ATWATER, W G and BENEDICT, F G *Carnegie Inst* ,
 Washington, Pub No 42, 1905
 BENEDICT, F G *Portable respiration apparatus for*
clinical use *Boston Med Surg J* , 1918, 178,
 667 Notes on the use of the portable respiration
 apparatus Ibid , 1920, 182, 243 Energy require-
 ment of intense mental effort *Science*, 1930, 71,
 567 BENEDICT, F G and COLLINS, W E *Clinical*
apparatus for measuring basal metabolism *Boston*
Med Surg J , 1920, 183, 449 BENEDICT, F G
 and Fox, E L *A method for the determination of*
the energy values of foods and excreta *J Biol*
Chem , 1925, 66, 783

- CATHCART, E P and CUTHBERTSON, D P *J Physiol* ,
 1931, 72, 349
 CATHCART, E P and MARKOWITZ, J J *Physiol* , 1927,
 63, 309
 DOUGLAS, C G *J Physiol* , 1911, 42, xvii
 DuBOIS, D and DuBOIS, E F *Arch Int Med* , 1916,
 17, 863
 HALDANE, J J *Physiol* , 1892, 13, 419
 HENDERSON, Y and HAGGARD, H. W *Am J Physiol* ,
 1925, 72, 264
 MURLIN, J R and CARPENTER, T M *Arch Int*
Med , 1911, 7, 184
 ROTH, P R *Boston Med Surg J* , 1921, 184, 222

Monographs and Reviews

- BOOTHBY, W M and SANDIFORD, L *Basal metabolic*
rate determination Saunders, Philadelphia, 1920
 BOOTHBY, W M *Basal metabolism* *Physiol Rev* ,
 1924, 4, 69
 CARPENTER, T M *A comparison of methods for deter-*
mining the respiratory exchange of man *Carnegie*
Inst , Washington, Pub No 216, 1915
 DuBOIS, E F *Basal metabolism in health and disease*
 Lea, Philadelphia, 1927
 HARDING, V J *Metabolism in pregnancy* *Physiol*
Rev , 1925, 5, 279
 HARROP, G A *Diet in disease* Blakiston, Philadel-
 phia, 1930
 KING, J T *Basal metabolism* The Williams & Wil-
 kins Co , Baltimore, 1924
 KROGH, A *The respiratory exchange of animals and*
man Longmans, London, 1916
 LUSE, G *The elements of the science of nutrition*.
 Saunders, Philadelphia, 1928
 McCANN, W S *Calorimetry in medicine* *Medicine*,
 1924, 3, 1
 McLESTER, J S *Nutrition and diet in health and dis-*
ease Saunders, Philadelphia, 1931
 MURLIN, J R *Normal processes of energy metabolism*
In Barker's Endocrinology and Metabolism, vol
 3 Appleton, New York, 1922
 PETERS, J P and VAN SLYKE, D D *Quantitative*
clinical chemistry Vol. 1 The Williams & Wil-
 kins Co , Baltimore, 1932
 RICHARDSON, H B *The respiratory quotient* *Physiol*
Rev , 1929, 9, 61
 ROSE, M S *The foundations of nutrition* Macmillan,
 New York, 1933
 SHERMAN, H. C *Chemistry of food and nutrition*
 Macmillan, New York, 1933
 TALBOT, E B *Basal metabolism of children* *Physiol*
Rev , 1925, 5, 477

CHAPTER 46

PROTEIN METABOLISM

- ALMQUIST, H J and MECCHI, E J *Biol Chem* , 1940,
 135, 355

- BACH, S J Biological methylation Biol Rev, 1945, 20, 158
- BENEDICT, S R and OSTERBERG, E J Biol Chem, 1923, 66, 229
- BLOCH, K and SCHOENHEIMER, R J Biol Chem, 1940, 133, 633
- BOLLMAN, J L, MANN, F C and MAGATH, T B Effect of total removal of the liver on the formation of urea Am J Physiol, 1924, 69, 371 Effect of total removal of the liver on deamination Ibid, 1926, 78, 258 The effect of removal of the liver on the formation of ammonia Ibid, 1930, 92, 92
- BOOTHBY, W M Arch Int Med, 1934, 63, 39
- BORSOOK, H. and DUBNOFF, J W J Biol Chem, 1940, 132, 559
- BORSOOK, H. and KEIGHLEY, G L Proc Roy Soc, 1935, B 118, 488
- BORSOOK, H. and WINEGARDEN, H. M Proc. Nat Acad Sci, 1931, 17, 31 Ibid, p 75
- BORSOOK, H. and DUBNOFF, J W J Biol Chem, 1941, 141, 717
- BRAND, E and associates Am J Physiol, 1929, 90, 296 (Proc) J Biol Chem, 1930, 87, ix Ibid, 1932, 92, lxx J Am Med Assoc, 1933, 101, 1047
- BRAUNSTEIN, A E and KRITZMANN, M G Enzymologia, 1937, 2, 129
- CHAMBERS, W H. and LUSK, G J Biol Chem, 1930, 85, 611
- COHEN, P P Proc Fed Am Soc. Exp Biol, 1942, 1, 273
- COHEN, P P and HAYANO, M J Biol Chem, 1948, 166, 239, 251, Ibid, 172, 405
- COX, G J and ROSE, W C J Biol Chem, 1926, 68, 781
- CUSHING, H. and FULTON, M N Arch Int Med, 1932, 60, 649
- DANN, M. and CHAMBERS, W H. J Biol Chem, 1930, 89, 675
- DENIS, W. and associates Sulfates in blood J Biol Chem, 1921, 49, 311 The nonprotein sulfur of the blood in certain pathologic conditions. Arch Int Med, 1928, 41, 385
- DOCK, W Am J Physiol, 1931, 97, 117
- DU VIGNEAUD, V. and associates J Biol Chem, 1939, 131, 57
- EMBDEN, G. and SCHMITZ, E Biochem Z, 1910, 29, 423
- FOLIN, O Approximately complete analyses of thirty "normal" urines Am J Physiol, 1905, 13, 45 Laws governing the chemical composition of urine Ibid, p 66 A theory of protein metabolism. Ibid, p 117
- GAEBLER, O H. J Biol Chem, 1929, 81, 41
- GORNALL, A G. and HUNTER, A Biochem J, 1941, 35, 650
- GRAHAM, W R, HOUGHIN, O B. and TURNER, C W J Biol Chem, 1937, 120, 29
- HOLT, J E and associates Bull J Hopk Hosp, 1944, 74, 308
- HOUSSAY, B A. and BIASOTTI, A J Physiol, 1932, 77, 81 Endocrin, 1931, 15, 511
- KNOOP, F Z physiol Chem, 1910, 67, 489 Ibid, 1925, 146, 267
- KREBS, H A Biochem J, 1935, 29, 1951
- KREBS, H. A. and HENSELEIT, K Z physiol Chem, 1932, 210, 33
- LARSON, P S. and CHAIKOFF, I L J Nutrit, 1937, 13, 287
- LEUTHARDT, F Z physiol Chem, 1938, 252, 238
- LUNDGAARD, E Skand Arch Physiol, 1931, 62, 223 Ibid, p 243
- MACLEOD, J J R The fuel of life Princeton University Press, 1928
- MANN, F C. and associates Am J Physiol, 1928, 87, 497
- MARTIN, G J Proc Soc. Exp Biol Med, 1944, 55, 182
- NEEDHAM, D M Biochem J, 1930, 24, 208
- RABINOWITCH, I M J Biol Chem, 1929, 83, 333
- RAPPORT, D. and BEARD, H H J Biol Chem, 1927, 73, 299
- RATNER, S Fed Proc, 1949, 8, 603
- ROSE, W C. and BUNNEY, W E Growth upon diets practically devoid of arginine J Biol Chem, 1928, 76, 521 ROSE, W C. and SCULL, C W Arginine metabolism Ibid, 1930, 89, 109
- SADOU, D P Missouri Agric. Exp Station, Research Bull, 1947, 108, 1
- SCHOENHEIMER, R. and associates J Biol Chem, 1942, 144, 541
- SHERWIN, C P. and HARROW, B J Biol Chem, 1926, 70, 683
- TROWELL, O A J Physiol, 1941, 100, 432
- DU VIGNEAUD, V Harvey Lect 1942-43, Biol Symposia, 1941, 5, 234
- WALKER, M B Lancet, 1934, 1, 1200 Ibid, 1935, 1, 448

Monographs and Reviews

- BEARD, H. H. The biochemistry of creatine and creatinine. Ann Rev Biochem, 1941, 10, 246
- BORSOOK, H. The specific dynamic action of protein and amino acids in animals Biol Rev, 1936, 11, 147
- CATHCART, E P The physiology of protein metabolism Longmans, London, 1921
- DAKIN, H. D. Oxidations and reductions in the animal body Longmans, London, 1922
- HUNTER, A. Creatine and creatinine Physiol Rev, 1922, 2, 586 Creatine and creatinine Longmans, London, 1928
- KOTAKI, Y. The metabolism of proteins and amino acids Ann Rev Biochem, 1934, 3, 193
- LEWIS, H. B. Sulfur metabolism Physiol Rev, 1924, 4, 394

- LUSK, G The elements of the science of nutrition Saunders, Philadelphia, 1906
- LUSK, G Specific dynamic action of various food factors *Medicine*, 1922, 1, 311 *J Nutrition*, 1930, 3, 519
- MANN, F C The effects of complete and partial removal of the liver *Medicine*, 1927, 6, 419
- MITCHELL, H H and HAMILTON, T S The biochemistry of the amino acids Chemical Catalog Co, New York, 1929
- OSBORNE, F B The vegetable proteins Longmans, London, 1924
- PETERS, J P and VAN SLYKE, D D Quantitative clinical chemistry Vol I The Williams & Wilkins Co, Baltimore, 1931
- PLIMMER, R H A The chemical constitution of the proteins Longmans, London, 1917
- RAPPORT, D The interconversion of the major food-stuffs *Physiol Rev*, 1930, 10, 349
- SCHOENHEIMER, R and RITTENBERG, D The study of intermediary metabolism of animals with the aid of isotopes *Physiol Rev*, 1940, 20, 218
- SHERMAN, H C The chemistry of food and nutrition Macmillan, New York, 1933
- WILHELMJ, C M The specific dynamic action of food *Physiol Rev*, 1935, 15, 202
- DU VIGNEAUD, V and DYER The chemistry and metabolism of compounds of sulfur *Ann Rev Biochem*, 1936, 5, 159
- OSBORNE, T B and MENDEL, L B Series of papers in *Journal of Biological Chemistry* from 1912, Vol 12, to 1920, Vol 41 OSBORNE, T B, MENDEL, L B and FERRY, E L A method of expressing numerically the growth-promoting value of proteins *J Biol Chem*, 1919, 37, 223
- EPSTEIN, S S and ROSE, W C Unpublished data cited by Rose in *Physiol Rev*, 1938, 18, 109
- ROSE, W C and associates The relation of arginine and histidine to growth *J Biol Chem*, 1924, 61, 747 Growth upon diets practically devoid of arginine *Ibid*, 1928, 76, 521 Arginine metabolism *Ibid*, 1930, 89, 109 Feeding experiments with mixtures of highly purified amino acids I The inadequacy of diets containing nineteen amino acids *Ibid*, 1931, 94, 155 II The supplementing effect of proteins *Ibid*, p 167 III The supplementing effect of casein fractions *Ibid*, p 173
- WILLCOCK, E G and HOPKINS, F G J *Physiol*, 1906, 35, 88

Monographs and Reviews

- McCOLLUM, E V and SIMMONDS, N The newer knowledge of nutrition Macmillan, New York, 1925
- MENDEL, L B Nutrition and growth Harvey Lectures, 1914-15, 10, 101
- MENDEL, L B Nutrition the chemistry of life Yale University Press, 1923
- MITCHELL, H H The nutritive value of proteins *Physiol Rev*, 1924, 4, 424
- MITCHELL, H H and HAMILTON, T S The biochemistry of the amino acids Chemical Catalog Co, New York, 1929
- ROSE, W C The significance of the amino acids in nutrition Harvey Lectures, 1934-35, 30, 49 The nutritive significance of the amino-acids *Physiol Rev*, 1938, 18, 109

CHAPTER 47

PROTEIN METABOLISM (*Continued*)

- BLOCK, R J and MITCHELL, H H *Nutrit Abst and Rev*, 1946, 16, 249
- BOOTHBY, W M, SANDIFORD, I, SANDIFORD, K and SLOSSE, J The effect of thyroxin on the respiratory and nitrogenous metabolism of normal and myxedematous subjects I A method of studying the reserve or deposit protein, with a preliminary report of the results obtained *Trans Assoc. Am Physicians*, 1925, 40, 195
- DANN, W J and DARBY, W J Appraisal of nutritive state (nutrition) in humans *Physiol Rev*, 1945, 25, 326
- HOLT, L E *Proc. Am Fed Biol*, 1942, 1, 117
- HOLT and colleagues See HOLT and NAJJAR.
- HOLT, L E and NAJJAR, V A *Proc. Fed Am Soc. Exp Biol*, 1942, 1, 117
- MADDEN, S C and associates *J Exp Med*, 1939, 69, 721
- McCOLLUM, E V and DAVIS, M The influence of the plane of protein intake on growth *J Biol Chem*, 1915, 20, 415 McCOLLUM, E V, SIMMONDS, N and PARSONS, H T Supplementary relationship between the proteins of certain seeds *J Biol Chem*, 1919, 37, 155 Supplementary protein values in foods *Ibid*, 1921, 47, 111 175 207 255

CHAPTER 48

THE METABOLISM OF THE PURINE AND PYRIMIDINE BASES

- BAUER, W and KLEMPERER, F *New Eng J Med*, 1944, 231, 681
- BENEDICT, S R., DAVIS, A R and NEWTONE, E B The combined uric acid in beef blood *J Biol Chem*, 1922, 54, 595
- FOLIN, O, BERGLUND, H. and DERICK, C J *Biol Chem*, 1924, 60, 361
- KREBS, H A and associates *Biochem J*, 1936, 30, 1380
- MANN, F C and associates *Am J Physiol*, 1925, 72, 629

Monographs and Reviews

- BENEDICT, S R Uric acid in its relations to metabolism Harvey Lectures, 1915-16, 11, 346

- DAKIN, H. D. Oxidations and reductions in the animal body Longmans, London, 1912 Physiological oxidations *Physiol Rev*, 1921, 1, 394
- DAVIDSON, J. N. The Biochemistry of the Nucleic Acids Methuen, London, 1950
- GARROD, A. E. Inborn errors of metabolism Frowde, London, 1923
- GREENSTEIN, J. P. Nucleoproteins Advances in protein chemistry 1944, 1, 209
- JONES, W. Nucleic acids, their chemical properties and physiological conduct Longmans, London, 1920
- LEVENE, P. A. and BASS, L. W. Nucleic acids Chemical Catalog Co., New York, 1931
- LORING, H. G. Chairman, Symposium on chemistry and metabolism of nucleic acids and their constituents *Proc Fed Am Soc Exper Biol*, 1947, 6, 487
- PETERS, J. P. and VAN SLYKE, D. D. Quantitative Chemical Chemistry, Vol I The Williams & Wilkins Co., Baltimore, 1932
- ROSE, W. C. Purine metabolism *Physiol Rev*, 1923, 3, 544
- CHAPTER 49
CARBOHYDRATE METABOLISM
- ABEL, J. J., GEILING, E. M. K., ROULLER, C. A., BELL, F. K. and WINTERSTEINER, O. *J Pharm Exp Therap*, 1927, 31, 65
- ALLEN, F. M. *J Metab Res*, 1922, 1, 75
- ANDERSON, E., RIOCH, O. M. and HAYMAKER, W. *Acta Neurovegetativa, Sonderabdr aus Bd V*, Heft 1-2, 1952
- BANTING, F. G. and BEST, C. H. *J Lab Clin Med*, 1922, 7, 251
- BERGMAN, H. and DRURY, D. R. *Am J Physiol*, 1938, 124, 279
- BERNARD, C. *Compt rend soc. biol*, 1849, 1, 14
- BLUM, F. *Deut Arch klin Med*, 1901, 71, 146
- BORNSTEIN, J. *Diabetes*, 1953, 2, 23
- BRITTON, S. W. and SILVETTE, H. *Am J Physiol*, 1932, 100, 693
- BROWNE, J. S. L. Josiah Macy Jr Foundation Conference, 1943
- BULATAO, E. and CARLSON, A. J. *Am J Physiol*, 1924, 69, 107
- BÜRGER, M. *Zeit f ges inn Med*, 1947, 2, 311
- CAMPBELL, J. and BEST, C. H. *Lancet*, 1938, 1, 1444
- CANNON, W. B., SHOHL, A. T. and WRIGHT, W. S. *Am J Physiol*, 1911, 29, 280
- CONN, J. W., LOUIS, L. H. and JOHNSTON, M. W. *Proc. Am Diabetes Assoc*, 1948, 8, 213
- CORI, G. T. and CORI, C. F. *J Biol Chem*, 1927, 72, 615
- CRANDALL, L. A. and CHERRY, I. S. *Am J Physiol*, 1939, 126, 658
- CRANOALL, L. A. and LIPSCOMB, A. *Am J Physiol*, 1947, 148, 312
- DANN, M. and CHAMBERS, W. H. *J Biol Chem*, 1933, 100, 493
- DE MAYER, J. *Arch Fisiol*, 1909, 7, 96
- DICKSON, W. H. and WILSON, M. J. *J Pharm Exp Therap*, 1924, 24, 33
- DOHAN, F. C. and LUCKENS, F. D. *Am J Physiol*, 1939, 125, 188
- DONHOFFER, C. and MACLEOD, J. J. R. *Proc Roy Soc*, 1932, B 110, 125
- DUNN, J. S., SHEEHAN, H. L. and MCLEITCHIE, N. G. B. *Lancet*, 1943, 1, 484
- EVANS, H. M., MEYER, K., SIMPSON, M. E. and REICHERT, F. L. *Proc Soc. Exp Biol Med*, 1932, 29, 857
- GELLHORN, E., FELDMAN, J. and ALLEN, A. *Endocrin*, 1941, 29, 137
- GEMMILL, C. L. *Bull J Hopk Hosp*, 1941, 68, 329
- GRIFFITHS, J. P. and WATERS, E. T. *Am J Physiol*, 1936, 117, 134
- GRIFFITHS, M. J. *J Biol Chem*, 1948, 172, 853
- GROSSMAN, M. I. and STEIN, I. F. *J Applied Physiol*, 1948, 1, 263
- HAGEDORN, H. C., JENSEN, B. N., KRARUP, N. B. and WODSTRUP, I. *J Am Med Assoc*, 1936, 106, 177
- HAIST, R. E., CAMPBELL, J. and BEST, C. H. *New Eng J Med*, 1940, 223, 607
- HALLAS MOLLER, K., JERSILD, M., PETERSEN, K. and SCHLICHTKRULL, J. J. A. M. A., 1952, 150, 1667
- HAM, A. W. and HAIST, R. E. *Am J Pathol*, 1941, 17, 787
- HARRIS, S. J. *Am Med Assoc*, 1924, 83, 729
- HÉDON, E. *J Physiol Path gen*, 1927, 25, 1
- HIMSWORTH, H. P. *Clinical Sci*, 1935, 2, 67
- HOFMEISTER, F. *Arch exper Path Pharmacol*, 1889, 25, 240
- HOUSSAY, B. A. *Endocrin*, 1944, 35, 158
- HOUSSAY, B. A. and BIASOTTI, A. *Rev soc argent. biol*, 1930, 6, 8
- HOUSSAY, B. A., FOGLIA, V. G., SMYTH, F. S., RIETTI, C. T. and HOUSSAY, A. B. *J Exp Med*, 1942, 75, 547
- HOUSSAY, B. A. and POTICK, D. *Compt. rend soc biol*, 1929, 101, 940
- JACOBS, H. R. *Proc. Soc. Exp Biol Med*, 1937, 37, 407
- JENSEN, H. and GRATTAN, J. F. *Am J Physiol*, 1940, 128, 270
- KERR, R. B. and BEST, C. H. *Am J Med Sci*, 1937, 194, 149
- KRAHL, M. E. *Ann N Y Acad Sci*, 1951, 54, 694
- LAZAROW, A. *Diabetes*, 1952, 1, 171
- LEHMANN, W. L. *Arch exper Path Pharmacol*, 1874, 2, 463
- LONG, C. N. H., KATZIN, B. and FRY, E. G. *Endocrin*, 1940, 26, 309
- LONG, C. N. H. and LUKENS, F. D. W. *J Exp Med*, 1936, 63, 465

- LUKENS, F D W and DOHAN, F C *Endocrin*, 1938, 22, 51
- LUKENS, F D W and DOHAN, F C *Endocrin*, 1942, 30, 175
- LUNDBAEK, K and GORANSON, E S *Nature*, 1948, 162, 1002
- MANF, F C and MAGATH, T B *Arch Int Med*, 1922, 30, 73
- MERING, J VON *Verhandl Congresses innere Med*, 1887, 6, 349
- MERING, J VON and MINKOWSKI, O *Arch exper Path Pharmacol*, 1889, 26, 371
- MIRSKY, I A and BROH-KAHN, R H *Arch Biochem*, 1949, 20, 1 10
- OTTAWAY, J H. and SMITH, R H. Quoted by F G YOUNG, *Lancet*, 1948, 2, 955
- REINECKE, R M *Am J Physiol*, 1943, 140, 276
- RICHARDSON, K. C and YOUNG, F G *J Physiol*, 1937, 91, 352
- RICHARDSON, K C and YOUNG, F G *Lancet*, 1938, 1, 1098
- SCOTT, D A. *Biochem J*, 1934, 28, 1592
- SCOTT, D A and FISHER, A M *J Pharm Exp Therap*, 1936, 58, 78
- SOSKIN, S, ALLWEISS, M D and COHN, D J *Am J Physiol*, 1934, 109, 155
- STADIE, W C *J Biol Chem*, 1947, 170, 55
- STADIE, W C, HAUGAARD, N and VAUGHAN, M J *Biol Chem*, 1952, 199, 729
- STETTEN, DEW *Bull N Y Acad Med*, 1953, 29, 446
- SUTHERLAND, E W and CORI, C F *J Biol Chem*, 1948, 172, 737
- WICK, A N, DRURY, D R, BANCROFT, R. W and MACKAY, E M *J Biol Chem*, 1951, 188, 241
- WILDER, R M *Diabetes*, 1952, 1, 183
- WILDER, R M, ALLAN, F N, POWER, M H and ROBERTSON, H E *J Am Med Assoc*, 1927, 89, 348
- YOUNG, F G *Lancet*, 1936, 1, 237, 297 *Ibid*, 1937, 2, 372
- Monographs and Reviews*
- BERNARD, C *Leçons de physiologie expérimentale* Baillière, Paris, 1858
- BRODY, S and KIBLER, H. H. *Growth and development, with special reference to domestic animals* LII Relation between organ weight and body weight in growing and mature animals University of Missouri College of Agriculture, Agricultural Experiment Station, Research Bulletin 328, 1941
- CHAMBERS, W H. *Undernutrition and carbohydrate metabolism* *Physiol Rev*, 1938, 18, 248
- CORI, C F *Phosphorylation of glycogen and glucose* *Biol. Symposia*, 1941, 5, 131
- CREVELD, S VAN *Glycogen disease. Medicine*, 1939, 18, 1
- HAIST, R. F. *Factors affecting the insulin content of*
- HAWORTH, W N *The constitution of sugars* Arnold, London, 1929
- Hormonal factors in carbohydrate metabolism *Ciba Foundation Colloquia on Endocrinology*, 6, 1953, for articles on enzymes and hormones in carbohydrate metabolism by G and C F CORI, F DICKENS, C DE DUVE, C N H LONG, F D W LUKENS, A E WILHELM, F G YOUNG and others
- HOUSAY, B A and BIASOTTI, A. *The hypophysis, carbohydrate metabolism and diabetes* *Endocrin*, 1931, 15, 511
- JENSEN, H F *Insulin its chemistry and physiology* Oxford University Press, London, 1938
- KRARUP, N B *Clinical investigation into the action of protamine insulinate* Copenhagen, 1935
- LUSK, G *The elements of the science of nutrition* Saunders, Philadelphia, 1928
- MACLEOD, J J R. *Carbohydrate metabolism and insulin* Longmans, London, 1926
- MACLEOD, J J R *The fuel of life*. Princeton University Press, 1928
- XIX International Physiological Congress Symposium on Insulin, 1953 Recent progress in hormone research *Proc Laurentian Hormone Conference*, 1953, 8, for articles on insulin and growth hormone by F G YOUNG and R C DEBODO
- SAKEL, M *Neue Behandlungsmethode der Schizophrenie* Perles, Wien, 1935
- Third International Standard for Insulin MILES, A A., MUSSETT, M V and PERRY, W L M *Bull World Hlth Org*, 1952, 7, 445
- CHAPTER 50
- FAT METABOLISM
- ALLAN, F N, BOWIE, D J, MACLEOD, J J R and ROBINSON, W L *Brit J Exp Path*, 1924, 5, 75
- ANDERSON, W E and MENDEL, L B *J Biol Chem*, 1928, 76, 729
- ARTOM, C *Nature*, 1953, 171, 347
- BENNETT, M A, MEDES, G and TOENNIES, G *Growth*, 1944, 8, 59
- BEST, C H. and CAMPBELL, J J *Physiol*, 1936, 86, 190
- BEST, C H, HERSHEY, J M and HUNTSMAN, M E *J Physiol*, 1932, 75, 56
- BLOCH, K J *Biol Chem*, 1945, 157, 661
- BLOOR, W R, OKEY, R and CORNER, G W *J Biol Chem*, 1930, 86, 291
- BLUMBERG, H. and MCCOLLUM, E V *Science*, 1941, 93, 598
- BOYD, E M *J Biol Chem*, 1942, 143, 131
- BREUSCH, F L *Science*, 1943, 97, 490
- BUCHANAN, J M, SAKAMI, W, GURIN, S and WILSON, D W *J Biol Chem*, 1945, 159, 695
- BURR, G O and BURR, M M *J Biol Chem*, 1930, 86, 587
- CAMINADE, R. A., MAYER, A and VALLÉE, H *Ann*

- CHAIKOFF, I L, ENTENMAN, C and MONTGOMERY, M L J Biol Chem, 1947, 168, 177
- CHAIKOFF, I L and SOSKIN, S Am J Physiol, 1928, 87, 58
- DAFT, F S, SEBRELL, W H. and LILLIE, R D Proc. Soc. Exp Biol Med, 1942, 50, 1
- DU VIGNEAUD, V, CHANDLER, J P, MOYER, A W and KEPPEL, D M J Biol Chem, 1939, 131, 57
- DYE, J A and CHIDSEY, J L Am J Physiol, 1939, 127, 745
- EMBDEN, G and KALBERLAH, F Beitr Chem Physiol Path, 1906, 8, 121
- ENTENMAN, C, CHAIKOFF, I L and ZILVERSMIT, D B J Biol. Chem, 1946, 166, 15
- FISHER, N F Am J Physiol, 1924, 67, 634
- GARDNER, J A. and GANSBPOUGH, H. Biochem J, 1925, 19, 667
- GOLDMAN, D S, CHAIKOFF, I L, REINHARDT, W O, ENTENMAN, C and DAUBEN, W G J Biol Chem, 1950, 184, 727
- GYÖRGY, P and GOLDBLATT, H. J Exp Med, 1942, 75, 355
- HARTROFT, W S and BEST, C H Brit Med J, 1949, 1, 423
- HARTROFT, W S and RIDOUT, J H. Am J Path, 1951, 27, 951
- HETRIQUES, V and HANSEN, C Skand Arch Physiol, 1901, 11, 151
- HERSHEY, J M and SOSKIN, S Am J Physiol, 1931, 98, 74.
- HUNTER, F E and LELOIR, L F J Biol Chem, 1945, 159, 295
- HURLEY, W H. Quart J Med, 1916, 9, 301
- JOWETT, M and QUASTEL, J H Biochem J, 1935, 29, 2181
- KNOOP, F Beitr Chem. Physiol Path., 1905, 6, 150
- LEACH, E H. J Physiol, 1938, 93, 1
- LYNEN, F Federation Proc., 1953, 12, 683
- MCAMIS, A J, ANDERSON, W E and MENDEL, L B J Biol Chem, 1929, 82, 247
- MCCARTHER, C S, LUCAS, C C and BEST, C H. Biochem J, 1947, 41, 612
- MACKAY, E M J Clin Endocrin, 1943, 3, 101
- MACKAY, E M, BARNES, R H., CARNE, H. O and WICK, A. N J Biol Chem 1940, 135, 157
- MACKAY, E M, WICK, A. N and BARNUM, C P J Biol Chem, 1940, 136, 503
- MAHLER, H. R. Federation Proc., 1953, 12, 694.
- MEDES, G Acta Med Skand, 1947, Supp 196, 646
- MEDES, G, FLOYD, N F and WEINHOUSE, S J Biol Chem, 1946, 162, 1
- MEDES, G, WEINHOUSE, S and FLOYD, N F J Biol Chem., 1945, 157, 35
- MILLER, L L and WHIPPLE, G H. J Exp Med, 1942, 76, 421
- MONTGOMERY, M L, ENTENMAN, C, GIBBS, G E and CHAIKOFF, I L Proc Soc. exp Biol. Med, 1940, 43, 349
- PERLMAN, I and CHAIKOFF, I L J Biol Chem, 1939, 127, 211
- PIHL, A. and BLOCH, K J Biol Chem, 1950, 183, 431
- QUASTEL, J H and WHEATLEY, A H M Biochem. J, 1933, 27, 1753
- SNAPPER, L, GRÜNBAUM, A and MENDES DE LEON, C Biochem Zeit, 1928, 201, 473
- SNAPPER, I, GRÜNBAUM, A and NEUBERG, J Biochem Zeit, 1926, 167, 100
- STERN, J R and OCHOA, S J Biol Chem, 1952, 198, 313
- STETTIN, DEW, JR J Biol Chem, 1941, 140, 143
- TUCKER, H F and ECKSTEIN, H C J Biol Chem., 1937, 121, 479
- VERZÁR, F and KÚTH, A von Biochem Zeit, 1931, 230, 451
- WEINHOUSE, S, MEDES, G, and FLOYD, N F J Biol Chem, 1944, 155, 143
- WEINMAN, E O and CHAIKOFF, I L, ENTENMAN, C and DAUBEN, W G J Biol Chem, 1950, 187, 643
- WELCH, A D and LANDAU, R L J Biol Chem, 1942, 144, 581

Monographs and Reviews

- BLOCH, K Intermediary metabolism of cholesterol Circulation, 1950, 1, 214
- BLOCH, K Interrelationships of lipid and carbohydrate metabolism Ann Rev Biochem, 1952, 21, 273
- BLOOM, W R. Biochemistry of the Fatty Acids Reinhold Publ Corp, N Y, 1943
- BULL, H B The Biochemistry of the Lipids John Wiley & Sons, Inc., N Y, 1937
- COOK, R P Cholesterol metabolism Nutr Abs and Rev, 1942, 12, 1
- DAKIN, H. D Oxidations and Reductions in the Animal Body Longmans, Green and Company, London, 1922
- DEUEL, H. J, JR. The Lipids Their Chemistry and Biochemistry, Vol 1 Interscience Publishers, N Y, 1951
- DORFMAN, R I and UNGAR, F Metabolism of Steroid Hormones. Burgess Publ Co, Minneapolis, Minn, 1953
- DRILL, V A Hepatotoxic agents mechanisms of action and dietary interrelationship Pharmacol Rev, 1952, 4, 1
- DU VIGNEAUD, V A Trail of Research in Sulfur Chemistry and Metabolism and Related Fields Cornell Univ Press, Ithaca, N Y, 1952
- FRAZER, A. C Fat metabolism Ann Rev Biochem, 1952, 21, 245
- GOULD, R. G Lipid metabolism and atherosclerosis Am J Med, 1951, 11, 209
- GRIFFITH, W H. The relation of choline to the kidneys Biol Symposia, 1941, 5, 193
- GURIN, S and BRADY, R. O The in vitro synthesis of

- lipids and its hormonal control Recent progress in hormone research, 1953, 8, 571
- HEVESY, G Application of radioactive indicators in biology *Ann Rev Biochem*, 1940, 9, 641
- KATZ, L N and STAMLER, J Experimental Atherosclerosis Charles C Thomas, Springfield, Ill, 1953
- KEYS, A Atherosclerosis a problem in newer public health. *J Mt Sinai Hospital*, 1953, 20, 118
- LEATHES, J B and RAPER, H S *The Fats* Longmans, London, 1925
- LIPMANN, F On chemistry and function of coenzyme A *Bacteriological Rev*, 1953, 17, 1
- LONG, C N H The relation of cholesterol and ascorbic acid to the secretion of the adrenal cortex *Recent Progress in Hormone Research*, 1947, 1, 99
- LONGENECKER, H E The formation of animal body fat *Biol Symposia*, 1941, 5, 99
- LUCAS, C C and BEST, C H *Vitamins and Hormones*, 1, 1942, Academic Press Inc, New York
- SCHOENHEIMER, R *The Dynamic State of Body Constituents* Harvard Univ Press, Cambridge, Mass, 1942
- SCHOENHEIMER, R and RITTENBERG, D The application of isotopes to the study of intermediary metabolism *Science*, 1938, 87, 221
- SINCLAIR, R G Physiology of the phospholipids *Physiol Rev*, 1934, 14, 351
- SINCLAIR, R G The anabolism and function of the phospholipids *Biol Symposia*, 1941, 5, 82
- SOSKIN, S and LEVINE, R The origin and regulation of ketone bodies from fats *Biol Symposia*, 1941, 5, 64
- STRAIN, W H The steroids Chapter 19 in *Organic Chemistry*, ed by H GILMAN Wiley and Sons, Inc, N Y, 1943
- TURNER, R H Serum Proteins and Lipids Josiah Macy, Jr Foundation, Trans 10th Conf Liver Injury, New York, 1951, 13
- WERTHEIMER, E and SHAPIRO, B The physiology of adipose tissue *Physiol Rev*, 1948, 28, 451
- WITTCOFF, H *The Phosphatides* Reinhold Publ Corp, N Y, 1951
- CAMPBELL, W R Dietary factors in health and disease *Can Med Assoc J*, 1935, 33, 379
- CANNON, P R, *Sci monthly*, 1943, 56, 5, *Medicine and the War*, W H Taliaferro, Ed, Univ Chicago Press, 1944, *J Am Diet Assoc*, 1944, 20, 77, *J Am Med Assoc*, 1945, 120, 360
- CATHCART, E P *The physiology of protein metabolism* Longmans, London, 1912
- DRUMMOND, SIR J, *Roy Soc Arts*, 1946, 94, 470, quoted in *The Biology of Human Starvation*, Univ Minnesota Press, A Keys, Ed, 1950
- HARROP, G A *Diet in disease* Blakiston, Philadelphia, 1930
- KEYS, A and assoc *The Biology of Human Starvation* University of Minneapolis, Minnesota Press, 1950
- LUCK, G *The elements of the science of nutrition* Saunders, Philadelphia, 1928 *The physiological effect of undernutrition* *Physiol Rev*, 1921, 1, 523
- McKAY, C M *Transactions State College Physicians Phila*, 1942, 10, 1
- MORGULIS, S *Fasting and undernutrition* Dutton, New York, 1923

CHAPTER 52

Snath

METABOLISM IN MUSCULAR ACTIVITY

- ANDERSON, R J and LUSK, G *Proc Nat Acad Sci*, 1917, 3, 386
- ASTBURY, W T *Ann Rev Bioch*, 1939, 8, 113
- BAILEY, K *Biochem J*, 1942, 36, 121
- BENEDICT, F G and CATHCART, E P *Carnegie Inst Washington, Pub No 187*, 1913
- BEST, C H, FURUSAWA, K and RIDOUT, J H *Proc Roy Soc.*, 1929, B 104, 119
- CANZANELLI, A and RAPPORT, D *Am Heart J*, 1932, 102, 325
- CARPENTER, T M and FOX, E L *Am J Physiol*, 1929, 90, 307
- EGGLETON, P and EGGLETON, G P *Biochem J*, 1927, 21, 190 *J Physiol*, 1927, 63, 155 *Ibid*, 1928, 65, 15
- ENGLEHARDT, W A. Recent advances in enzymology, 1946, 6, 147
- ENGLEHARDT, W A and LJUBIMOWA, M N *Nature*, 1939, 144, 668
- FISKE, C H and SUBBAROW, Y *Science*, 1927, 65, 401 *Ibid*, 1928, 67, 169 *J Biol Chem*, 1929, 81, 629
- FLETCHER, W M and HOPKINS, F G *J Physiol*, 1907, 35, 247 *Proc Roy Soc*, 1917, B 89, 444
- FLOCK, E V, INGLE, D J and BOLLMAN, J L *J Biol Chem*, 1939, 129, 99
- GEMMIL, C L Fuel of muscular exercise *Phys Rev*, 1942, 22, 32
- HARTREE, W J *Physiol*, 1931, 72, 1 *Ibid*, 1932, 75, 273
- HILL, A V, LONG, C N H and LUPTON, H. *Proc*
- CHAPTER 51
- METABOLISM IN STARVATION, SEMISTARVATION AND OBESITY
- ADDIS, T, POO, L J and LEW, W J *Biol Chem*, 1936, 115, 111
- FOLIN, O and DENIS, W J *Biol Chem*, 1915, 21, 183
- HOWE, P E, MATTILL, H A and HAWK, P B *J Biol Chem*, 1912, 11, 103
- NEWBURGH, L H Obesity *Physiol Rev*, 1944, 24, 18
- Monographs and Reviews*
- BENEDICT, F G A study of prolonged fasting *Carnegie Inst, Washington, Pub 203*, 1915

- Roy Soc, 1924, B 97, 84 HILL, A. V and LUTTON, H Quart J Med, 1923, 16, 135
- HIMWICH, H E and ROSE, M I Am J Physiol, 1927, 81, 485
- KROGH, A. and LINDHARD, J Biochem J, 1920, 14, 290
- LOHMANN, K Ann Rev Biochem, 1938, 7, 125
- LUNDGAARD, E Untersuchungen über Muskelkontraktionen ohne Milchsäurebildung Biochem Z, 1930, 217, 162 Ibid, 227, 51 The biochemistry of muscle Ann Rev Biochem, 1938, 7, 377
- MEYERHOFF, O Nature, 1933, 132, 337, 373
- NEEDHAM, J and associates Nature, 1941, 147, 766
- SACKS, J Am J Physiol, 1940, 129, 227
- SACKS, J and SACKS, W C Am J Physiol, 1933, 105, 151
- SZENT GYÖRGYI, A Acta Physiol Skand, 1945, 9, suppl, 25 Nature of Life, a study of muscle Acad Press, New York, 1948
- Monographs and Reviews*
- BAINBRIDGE, F A The physiology of muscular exercise Third edition, rewritten by BOCK, A V and DILL, D B Longmans, London, 1931
- CARPENTER, T M Editorial review The fuel of muscular activity of man J Nutrit, 1931, 4, 281 Editorial review The effect of muscular exercise on the metabolism of ethyl alcohol J Nutrit, 1933, 6, 205 Carneg Inst Wash Suppl, 1937, 31
- CATRACART, E P The influence of muscular work on protein metabolism Physiol Rev, 1925, 5, 225
- EGGLETON, P Recent progress in the process of muscular contraction Biol Rev, 1933, 8, 46 Muscular exercise Paul, Lond, 1936
- FENN, W O Contractility, Article (P 445) in Physical Chemistry of Cells and Tissues Blakiston, Phila, 1945
- FISCHER, E Muscle contraction Biol Symp, 1941, 3, 211
- GEMMILL, C L The fuel of muscular energy Physiol Rev, 1942, 22, 32
- HILL, A V The mechanism of muscular contraction Physiol Rev, 1922, 2, 310 Muscular activity The Williams & Wilkins Co, Baltimore, 1926 Muscular movement in man. McGraw-Hill, New York, 1927 Adventures in biophysics Oxford University Press, 1931 The revolution in muscle physiology Physiol Rev, 1932, 12, 56
- MELLANBY, E Alcohol its absorption and disappearance from the blood under different conditions Med Res Com, London, Spec. Rep No 31, 1919
- MEYERHOFF, O Die chemischen Vorgänge im Muskel Springer, Berlin, 1930 The chemistry of anaerobic recovery of muscle New Eng J Med, 1939, 220, 49
- SACKS, J Changing concepts of the chemistry of muscular contraction Physiol Rev, 1941, 21, 217
- SCHNEIDER, E C The Physiology of Muscular Activity, 4th Ed, revised by P V Karpovitch Saunders, Phila and London
- WOOD, H G The fixation of carbon dioxide and the inter-relationships of the tricarboxylic acid cycle Phys Rev, 1946, 26, 198
- CHAPTER 53
- THE BODY TEMPERATURE HEAT BALANCE
- BARBOUR, H G and PRINCE, A L J Pharm Exp Therap, 1914, 6, 1
- BARCROFT, J and IZQUIERDO, J J The effect of temperature on the frequency of heart and respiration in the guinea pig and cat J Physiol, 1931, 71, 364 See also, DWORKIN, S Observations on the central control of shivering Am J Physiol, 1930, 93, 227
- BAZETT, H C, ALPERS, B J and ERB, W H Arch Neur Psychiat, 1933, 30, 729
- CLARK, G, MAGOUN, H W and RANSON, S W J Neurophysiol, 1939, 2, 61 202
- COON, J M and ROTHMAN, S J Pharm Exp Therap, 1941, 73, 1
- BURN, J H J Physiol, 1925, 60, 365
- DUBOIS, E F J Am Med Assoc, 1921, 77, 352
- HARDY, J D and DUBOIS, E F Nat Acad Sci, 1940, 26, 389
- HARDY, J D and MUSCHENHEIM, C J Clin Invest, 1934, 13, 817 Ibid, 1936, 15, 1
- KELLER, A D J Neurophysiol, 1938, 1, 543
- KELLER, A D and HARE, W K Heat regulation in medullary and mid brain preparations Proc Soc Exp Biol Med, 1931, 29, 1067 The hypothalamus and heat regulation Ibid, p 1069
- KUNO, Y Significance of sweating in man Lancet, 1930, 2, 912
- MOSS, K V Proc Roy Soc, 1923, B 95, 181
- RANSON, S W and associates J Neurophysiol, 1939, 2, 61
- SHAFFER, P A Metabolism in typhoid fever J Am Med Assoc, 1908, 51, 974 SHAFFER, P A and COLEMAN, W Protein metabolism in typhoid fever Arch Int Med, 1909, 4, 538 COLEMAN, W Diet in typhoid fever J Am Med Assoc., 1909, 53, 1145 Five years' experience with high calory diet in typhoid Ibid, 1912, 59, 363
- UOTILA, U U Endocrin, 1939, 25, 63
- WATERLOW, J C Brit Med Bull, 1947, 5, 1012
- Monographs and Reviews*
- BARBOUR, H G The heat-regulating mechanism of the body Physiol Rev, 1921, 1, 295
- BAZETT, H. C Physiological responses to heat. Physiol Rev, 1927, 7, 531
- BURTON, A. C Clothing and heat exchange Proc Fed Am Soc. Exp Biol., 1946, 5, 344
- DEIGHTON, T Physical factors in body temperature maintenance and heat elimination Physiol Rev, 1933, 13, 427

- DuBois, E F Metabolism in fever and certain infections In *Barker's Endocrinology and Metabolism*, 1922, 4, 95 Heat loss from human body *Bull N Y Acad Med*, 1939, 15, 143, Fever and the Regulation of Body Temperature. C C Thomas, Springfield, Ill, 1948
- HILL, L and CAMPBELL, A Health and environment Arnold, London, 1925
- KUNO, Y The physiology of human perspiration Churchill, Lond, 1934
- LUSK, G The elements of the science of nutrition Saunders, Philadelphia, 1906
- MACCALLUM, W G Fever Harvey Lectures, 1908-09, 4, 27
- MACHLE, W and HATCH, T F Heat, man's exchanges and physiological responses *Phys Rev*, 1947, 27, 200
- MARTIN, C Thermal adjustments of man and animals to external conditions *Lancet*, 1930, 2, 561 617 673
- OTT, I Fever, its thermotaxis and metabolism Hoeber, New York, 1915
- RANSON, S W Regulation of body temperature Assoc for Research in Nervous and Mental Disease 1940, 20, 342
- SUNDSTROM, E S The physiological effects of tropical climate *Physiol Rev*, 1927, 7, 320
- 1930, 179, 91 FLETCHER, A A J *Lab Clin Med*, 1930, 15, 3
- FRASER, H and STANTON, A T *Lancet*, 1909, 1, 451
- FUNK, C J *Physiol*, 1911, 43, 395
- GAVIN, G and MCHENRY, E W J *Biol Chem*, 1941 141, 619
- GOLDBERGER, J The relation of diet to pellagra J Am Med Assoc, 1922, 78, 1676 Pellagra, its nature and prevention U S Pub Health Rep, 1927, 42, 2193 GOLDBERGER, J and TANNER, W F A study of the treatment and prevention of pellagra U S Pub Health Reports, 1924, 39, 87 GOLDBERGER, J, WHEELER, G A, LILLIE, R D and ROGERS, L M A further study of butter, fresh beef and yeast, as pellagra preventives with consideration of the relation of factor P-P of pellagra (and black tongue of dogs) to vitamin B U S Pub Health Reports, 1926, 41, 297 GOLDBERGER, J, WHEELER, G A and SULLIVAN, M Experimental production of pellagra in human subjects by diet *Bull U S Hygienic Lab No* 120, 1920
- GRIFFITH, J Q and associates *Proc Soc Exp Biol. Med*, 1944, 55, 228
- HARRIS, L J, RAY, S N and WARD, A *Bioch J*, 1933, 27, 2011
- HARRIS, S A and FOLKERS, K J *Am Chem Soc*, 1939, 61, 1245
- HARRIS, S A and associates *Science*, 1943, 97, 447
- HERRIN, R C and NICHOLS, H J *Am J Physiol*, 1939, 125, 786
- HESS, A F, LEWIS, J M and BARENBERG, L H J *Am Med Assoc*, 1933, 101, 657
- HIRST, E L and associates *J Chem Soc, London*, 1933, p 1419
- HOPKINS, F G J *Physiol*, 1912, 44, 425
- HOLST, A and FRÖHLICH, T J *Hygiene*, 1907, 7, 635
- JANSEN, B C P and DONATH, W F *Chem Weekblad*, 1926, 23, 201
- KING, C G and WAUGH, W A *Science*, 1932, 75, 357 *J Biol Chem*, 1932, 97, 325
- KINNERSLEY, H. W and PETERS, T A The relation between the lactic acid content of the brain and the symptoms of opisthotonus in rice-fed pigeons *Biochem J*, 1929, 23, 1126 Brain localisation of lactic acidosis in avitaminosis B₁ and its relation to the origin of symptoms *Ibid*, 1930, 24, 711 See also, PETERS, R A and O'BRIEN, J R Vitamin B₂ deficiency in the rat's brain *J Physiol*, 1935, 85, 454
- KöGL, F and TÖNNIS, B *Zeit physiol Chem*, 1936, 242, 43
- LEPKOVSKY, S, WOOD, C and EVANS, H M *J Biol Chem*, 1930, 87, 239
- McCOLLUM, E V and DAVIS, M The necessity of certain lipins in the diet during growth *J Biol Chem*, 1913, 15, 167 Observations on the isolation of the substance in butter fat which exerts a

CHAPTER 54

THE VITAMINS

- stimulating influence on growth J Biol Chem, 1914, 19, 245
- MACKAY, H. M M Arch Dis Child, 1934, 9, 133
- MELLANBY, E The presence in food stuffs of substances having specific harmful effects under certain conditions J Physiol, 1926, 61, xxiv The experimental production and prevention of degeneration in the spinal cord Brain, 1931, 54, 247 Nutrition in relation to bone growth and the nervous system Proc. Roy Soc. B, 1944, 132, 28 Vitamin A and bone growth, the reversibility of vitamin A deficiency changes J Physiol, 1947, 105, 382
- MELLANBY, E and GREEN, H. N Brit Med J, 1929, 1, 984
- MENDEL, L B J Am Med Assoc, 1932, 98, 1981
- MILAS, N A Science, 1946, 103, 581
- MINOT, G R., STRAUS, M B and COBB, S New Eng J Med, 1933, 208, 1244.
- MITCHELL, H. S J Biol Chem, 1919, 40, 399
- MOORE, T Vitamin A and carotene I The association of vitamin A activity with carotene in the carrot root Biochem J, 1929, 23, 803 V The absence of the liver oil vitamin A from carotene VI The conversion of carotene to vitamin A *in vivo* Ibid, 1930, 24, 692 VII The distribution of vitamin A and carotene in the body of the rat Ibid, 1931, 25, 275 Ibid, 1940, 34, 1321
- OLCOTT, H. S and MCCANN, D C J Biol Chem, 1931, 94, 185
- ORR, J B and RICHARDS, M B Biochem J, 1934, 28, 1259
- OSBORNE, T B and MENDEL, L B The relation of growth to the chemical constituents of the diet. J Biol Chem, 1913, 15, 311 The influence of butter fat on growth. Ibid, 1913, 16, 423 Ophthalmia and diet J Am Med Assoc., 1921, 76, 905
- PETERS, R A. Biochem J, 1936, 30, 2206
- PETERS, R A and associates Maintenance nutrition in the adult pigeon and its relation to torulin (vitamin B₁) Biochem J, 1930, 24, 1832
- READER, V Biochem J, 1929, 23, 689
- ROSENHEIM, O and DRUMMOND, J C Biochem J, 1925, 19, 753
- SCARBOROUGH, H. Biochem J, 1939, 33, 1400
- SMITH, M I and HENDRICK, E G U S Pub Health Reports, 1926, 41, 201
- SPIES, T D, COOPER, C and BLANKENHORN, M A The use of nicotinic acid in the treatment of pellagra J Am Med Assoc., 1938, 110, 622
- SPIES, T D, HIGHTOWER, D P and HUBBARD, L H J Am Med Assoc, 1940, 115, 292
- SPIES, T D and ASHE, W F J Am Med Assoc., 1939, 112, 2414
- SZENT GYÖRGYI, A Biochem J, 1928, 22, 1387
- SZENT GYÖRGYI, A and associates Nature, 1936, 138, 27, 798, 1057
- TAKAKI, K Lancet, 1906, 1, 1369
- WHIPPLE, D V and CHURCH, C F Proc. Am Soc Biol
- WILDERS, E La cellule, 1901, 18, 313 (Quoted by Rosenberg)
- WILLIAMS, R. J and associates J Am Chem Soc., 1935, 55, 2912
- WILLIAMS, R R J Am Chem Soc, 1936, 58, 1063
- WILLIAMS, R R and WATERMAN, R L The composite nature of vitamin B Proc Soc Exp Biol Med, 1927, 25, 1 The tripartite nature of vitamin B J Biol Chem, 1928, 78, 311
- WOLBACH, S B Science, 1937, 86, 569
- WOLBACH, S B and BESSEY, O A Arch Path, 1941, 32, 689
- WOODS, D D Br J Exp Path, 1940, 21, 74
- ZILVA, S S Biochem J, 1937, 31, 915 Ibid, 1488

Monographs and Reviews

- AYKROYD, W R Pellagra Nutrition Abstracts and Reviews, 1933, 3, 337
- BESSY, O H and WOLBUCK, S B Vitamin A, physiology and pathology Special article. J Am Med Assoc, 1938, 110, 2072
- COWGILL, G R The vitamin B requirement Yale University Press, 1934
- COWARD, K H Biological standardization of the vitamins Baillière, Tindall and Cox, Lond, 1938
- ELVENJEM, C A The vitamin B complex J Am Med Assoc., 1948, 138, 960
- Handbook of Nutrition, published for Am Med Assoc. by Blakiston, Phila, 1951
- HERTZ, R Biotin and the avidin-biotin complex Physiol Rev, 1946, 26, 479
- KING, C G Vitamin C, ascorbic acid Physiol Rev, 1936, 16, 238 The physiology of vitamin C Special article J Am Med Assoc, 1938, 111, 1458
- LIND, J, Treatise on scurvy London, 1757 Quoted from Med Res Coun, London Spec Rep No 167, 1932
- Medical Research Council, London Vitamins a survey of present knowledge Spec Rep No 167, 1932
- MELLANBY, E Nutrition and disease The interaction of clinical and experimental work Oliver, Edinburgh, 1934
- MENDEL, L B and others The vitamins A symposium on the present status of the knowledge of the vitamins Am Med Assoc Public, 1932
- MUNSELL, H. E Vitamin B₁ Special article J Am Med Assoc., 1938, 111, 927
- ROBERTSON, E C The vitamins and resistance to infection Medicine, 1934, 13, 123
- ROSENBERG, H R Chemistry and physiology of the vitamins Interscience Pub, New York, 1945
- SEBRELL, W H The relation of sulfonamide drugs to

vitamin deficiencies Harvey Lectures, 1943-44, 39, 288

WILLIAMS, R. R. and SPIES, T. D. Vitamin B₁ (thiamin) and its use in medicine Macmillan, New York, 1938

CHAPTER 55

THE VITAMINS (*Continued*)

ADAMSTONE, F. B. J. Morph and Physiol, 1931, 52, 47

ALMQUIST, H. J. J. Biol Chem, 1937, 120, 635

ALMQUIST, H. J. and KLOSE, A. A. J. Am Chem Soc, 1939, 61, 2557

ALMQUIST, H. J. and STOCKSTAD, E. L. R. J. Nutrit, 1936, 12, 329

ANDERSON, R. J. and NEWMAN, M. S. J. Biol Chem, 1933, 101, 773 Ibid, 1934, 103, 197

BILLS, C. E. J. Biol Chem, 1927, 72, 751

BINKLEY, S. B. and associates J. Biol Chem, 1939, 130, 219

BRUCE, H. M. and CALLOW, R. K. Biochem J, 1934, 28, 512 518

DAM, H. Biochem Zeit, 1929, 216, 475 Nature, 1934, 133, 1909 DAM, H. and GLAVIND, J. Lancet, 1938, 1, 720

EVANS, H. M. and BISHOP, K. S. On the existence of a hitherto unrecognized dietary factor essential for reproduction Science, 1922, 56, 650 Relation of vitamin E to growth and vigor J. Nutrit, 1928, 1, 23

EVANS, H. M., EMERSON, G. A. and EMERSON, O. H. Science, 1938, 88, 38

FERNHOLTZ, E. J. Am Chem Soc., 1938, 60, 700

FIESER, L. F. and associates J. Am Chem Soc, 1939, 61, 2559

GOLDBLATT, H. and SOAMES, K. M. Biochem J, 1923, 17, 446

GREAVES, J. D. and SCHMIDT, C. L. A. Proc. Soc. Exp Biol Med, 1937, 37, 43

HANKE, M. T. J. Am Dental Assoc, 1930, 17, 957

HESS, A. F. and UNGER, L. J. J. Am Med Assoc, 1921, 77, 39

HESS, A. F. and WEINSTOCK, M. J. Biol Chem, 1924, 62, 301

HOU, H. C. Chinese J. Physiol, 1928, 2, 345 Ibid, 1929, 3, 171 Ibid, 1930, 4, 79 345 Ibid, 1931, 5, 11

HOULDSCHINSKY, K. Deut Med Woch, 1919, 45, 712

HUME, E. M. and SMITH, H. H. Biochem J, 1923, 17, 364

KARRER, P. and colleagues Nature, 1938, 141, 1057 League of Nations Report of the permanent commission on biological standardization Quart Bull Health Organization, Spec No, January, 1935

MATTILL, H. A. and CONKLIN, R. E. J. Biol Chem, 1923, 55, 443 MATTILL, H. A. and STONE, N. C. Ibid, 1923, 55, 443

MCCOLLUM, E. V. and associates The effect of cod liver oil administered to rats with experimental rickets J. Biol Chem, 1921, 45, 343 The production of rachitis and similar diseases in the rat by deficient diets Ibid, 1921, 45, 333 A delicate biological test for calcium-depositing substances Ibid, 1922, 51, 41 An experimental demonstration of the existence of a vitamin which promotes calcium deposition Ibid, 1922, 53, 293 The effect of starvation on the healing of rickets Bull Johns Hopkins Hosp, 1922, 33, 31

McKEE, R. W. and associates J. Am Chem Soc, 1939, 61, 1259

MELLANBY, E. Experimental rickets Med Res Coun, London, Spec. Rep No 61, 1921 The effect of cereals and their interaction with other factors of diet and environment in producing rickets Ibid, No 93, 1925 The presence in foodstuffs of substance having specific harmful effects under certain conditions J. Physiol, 1926, 61, xxiv

MELLANBY, M. Med Res Coun, London, Spec. Rep No 140, 1929 and No 153, 1930

PALM, T. A. The Practitioner, 1890, 45, 271 321

ROSENHEIM, O. and WEBSTER, T. A. Lancet, 1925, 1, 1025 Ibid, 1927, 1, 306 Biochem J, 1927, 21, 389

SCHOUR, I. The effects of dietary deficiencies upon the oral structures Physiol Rev, 1945, 25, 442

STEENBOCK, H. and BLACK, A. J. Biol Chem, 1924, 61, 405 Ibid, 1925, 64, 263

TISDALL, F. F. and BROWN, A. J. Am Med Assoc., 1929, 92, 860 Am J Dis Child, 1927, 34, 721

TISDALL, F. F. and associates J. Am Dent Assoc, 1933, 20, 193 Ibid, 1934, 21, 1349

WADDELL, J. J. Biol Chem, 1934, 105, 711

WATSON, E. M. Can Med Assoc J, 1936, 34, 134

WINDAUX, A. and HESS, A. F. Nachr ges Wiss Göttingen, 1926, 2, 175 (Quoted in Med Res Coun, London, Spec Rep No 167)

Monographs and Reviews

BILLS, C. E. Physiology of the sterols including vitamin D Physiol Rev, 1935, 15, 1 The chemistry of vitamin D Special article J. Am Med Assoc, 1938, 110, 2150

BLUNT, K. and COWAN, R. Ultra-violet light and vitamin D in nutrition Chicago University Press, 1930

BUTT, H. R. and SNELL, A. M. Vitamin K Saunders, Philadelphia, 1941

EVANS, H. M. Vitamin E J. Am Med Assoc, 1932, 99, 469

HANKE, M. T. Diet and Dental Health, Chicago University Press, 1933

HESS, A. F. Rickets including osteomalacia and tetany Lea, Philadelphia, 1929

LAURENS, H. The physiological effects of radiant energy Chemical Catalog Co, New York, 1933

- The physiological effects of radiation. *Physiol Rev*, 1928, 8, 1
- MATTHILL, H. A. Vitamin E. Special article. *J. Am. Med. Assoc.*, 1938, 110, 1831
- MELLANBY, M. Influence of diet on the structure of teeth. *Physiol Rev*, 1928, 8, 545
- REED, C. I., STRUCK, H. C. and STECK, L. E. *Vitamin D*. University Chicago Press, 1939
- TOVERUD, K. U., *Acta Paediat. Scand* 1948, 36, 257
- See also Monographs and Reviews in Chapter 54

CHAPTER 56

DIETARY REQUIREMENTS

- BEARD, H. H. *Am. J. Physiol.*, 1927, 82, 577
- BURR, G. O. and BURR, M. M. A new deficiency disease produced by the rigid exclusion of fat from the diet. *J. Biol. Chem.*, 1929, 82, 345. On the nature and rôle of the fatty acids essential in nutrition. *Ibid.*, 1930, 86, 587
- WESSON, L. G. and BURR, G. O. The metabolic rate and respiratory quotients of rats on a fat-deficient diet. *Ibid.*, 1931, 91, 535. See also MCANIS, A. J., ANDERSON, W. E. and MENDEL, L. B. Growth of rats on fat-free diets. *J. Biol. Chem.*, 1929, 82, 247
- CATHCART, E. P. and MURRAY, A. M. T. *Med. Res. Coun.*, London, Spec. Rep. No. 151, 1931. *Ibid.*, No. 165, 1932. *J. Nutrit.*, 1931, 3, 483
- CLEMENTS, F. W. *Med. J. Aust.*, 1934, 1, 520
- DRUMMOND, J. C. *Lane lectures*, Stanford University Publications, 1934, 3, 7
- EVANS, H. M. and LEPOVSEY, S. J. *Biol. Chem.*, 1932, 95, 1-3. *Ibid.*, 99, 231
- FERGUSON, H. P., LEESON, H. J. and MCHENRY, E. W. *Canad. J. Pub. Health*, 1944, 35, 66, 241
- KROGH, A. and LINHARD, J. *Biochem. J.*, 1920, 14, 290
- LIEB, C. W. *J. Am. Med. Assoc.*, 1929, 93, 20
- OER, J. B. and GILES, J. L. *Med. Res. Coun.*, London, Spec. Rep. No. 155, 1931.
- RIGGS, E. et al. *Canad. J. Pub. Health*, 1943, 34, 193
- SHERMA, H. C. and HAWLEY, E. J. *Biol. Chem.*, 1922, 53, 375
- SINCLAIR, R. G. The physiology of the phospholipids. *Physiol. Rev.*, 1934, 14, 351
- SMADAL, J. E. and FAER, L. E. *J. Exp. Med.*, 1937, 65, 527, 541
- THOMAS, W. A. *J. Am. Med. Assoc.*, 1927, 88, 1559
- See also BISCHOFF, F. The influence of diet on renal and blood vessel changes. *J. Nutrit.*, 1932, 5, 431

Monographs and Reviews

- ATTWATER, W. V. and BRYANT, A. P. The chemical composition of American food materials. U. S. Department of Agriculture, Bull. No. 28, 1906
- RAYLES, W. M. The physiology of food and economy in diet. Longmans, London, 1917
- CHITTENDEN, R. H. *Physiological economy in nutrition*. Stokes, New York, 1907

- HARROP, G. A. *Diet in disease*. Blakiston, Philadelphia, 1930
- HUTCHISON, R. and MOTTRAM, V. H. *Food and the principles of dietetics*. Wood, New York, 1933
- MCCAY, D. *The protein element in nutrition*. Arnold, London, 1912
- MCLESTER, J. S. *Nutrition and diet in health and disease*. Saunders, Philadelphia, 1928
- PEARL, R. *The nation's food. A statistical study of a physiological and social problem*. Saunders, Philadelphia, 1920
- ROSE, M. S. *The foundations of nutrition*. Macmillan, New York, 1933
- SHERMAN, H. C. *Chemistry of food and nutrition*. Macmillan, New York, 1933

CHAPTER 57

THE PITUITARY BODY

- ALLAN, H. and WILES, P. J. *Physiol.*, 1932, 75, 23
- ALLEN, B. M. *Anat. Rec.*, 1916, 11, 486
- ANDERSON, E. M. and EVANS, H. M. *Am. J. Physiol.*, 1937, 120, 597
- ANSELMI, K. J. and HOFFMAN, F. Das Fettstoffwechselhormon des Hypophysenvorderlappens I. Nachweis, Darstellung und Eigenschaften des Hormons. *Klin. Woch.*, 1931, 10, 2380. Die pankreatische Substanz aus dem Hypophysenvorderlappen. *Klin. Woch.*, 1933, 12, 1435. Über die parathyreotrope Wirkung von Hypophysenvorderlappenextracten. *Ibid.*, 1934, 13, 45
- ARON, M. *Rev. franc. Endocrin.*, 1929, 8, 472
- ASCHNER, B. *Pflüger's Arch.*, 1912, 146, 1
- BAILEY, P. and BREMER, F. *Arch. Int. Med.*, 1921, 28, 773
- BELL, G. H. and ROBSON, J. M. *J. Physiol.*, 1935, 84, 351
- BEST, C. H. and CAMPBELL, J. Anterior pituitary extracts and liver fat. *J. Physiol.*, 1936, 86, 190. Production of diabetes in dogs by anterior-pituitary extracts. *Lancet*, 1938, 1, 1444.
- BIESONNETTE, T. H. *Proc. Assoc. Res. Nerv. Ment. Dis.*, 1936, 17, 361. *Wilson Bull.*, 1937, 49, 241. *Endocrin.*, 1938, 22, 92.
- BLAKE, P. T., COLLIP, J. B. and THOMSON, D. L. The effect of anterior pituitary extracts on acetone body excretion in the rat. *J. Physiol.*, 1934, 82, 385
- BLOUNT, R. F. *Proc. Nat. Acad. Sci.*, 1930, 16, 222
- BUEN, J. P. *J. Physiol.*, 1923, 57, 318
- BUEN, J. H. and LING, H. W. *Quart. J. Pharm. Pharmacol.*, 1933, 6, 31
- CARTER, G. S. and associates. *Nature*, 1943, 151, 728
- COLLIP, J. B. and ANDERSON, E. M. The production of serum inhibitory to the thyrotropic hormone. *Lancet*, 1934, 1, 76. COLLIP and associates. The adrenotropic hormone of the anterior pituitary lobe. *Lancet*, 1933, 2, 347. COLLIP, J. B. Prepara-

- tion of a purified and highly potent extract of growth hormone of anterior pituitary lobe *Proc Soc. Exp Biol Med*, 1933, 30, 544 The production of an increase in metabolic rates of thyroidectomized rabbits by certain pituitary extracts *Trans Am Assoc in Study of Goiter*, 1937 The specific metabolic principle of the pituitary and its relation to the melanophore principle *Endocrin*, 1938, 23, 718 *Ibid*, 1939, 25, 318
- CORKILL, A B, MARKS, H P and WHITE, W E *J Physiol*, 1933, 80, 193
- CROOKE, A C *J Path and Bact*, 1935, 41, 339
- CUSHING, H and DAVIDOFF, L M *Rockefeller Inst Med Res*, New York, Monograph No 22, 1927
- CUSHING, H and GOETSCH, E *Am J Physiol*, 1910, 27, 60
- DALE, H H *Biochem J*, 1909, 4, 427
- DANDY, W E and REICHERT, F L *Bull J Hopk Hosp*, 1938, 62, 122
- DUDLEY, H W *J Pharm Exp Therap*, 1919, 14, 295
- VAN DYKE, H B and assoc *J Pharmacol*, 1942, 74, 190
- EVANS, H M and LONG, J A *Anat Rec*, 1921, 21, 62 (*Proc*)
- EVANS, H M, MEYER, K and SIMPSON, M E *Memoirs Univ Calif*, 1933, 11, 1
- FERGUSON, J K W *Am J Physiol*, 1939, 126, 489
- FISHER, C, INGRAM, W R and RANSON, S W *Arch Neurol Psychiat*, 1935, 34, 124 *Anat Rec*, 1935, 63, 29 *Arch Int Med*, 1936, 57, 1067
- INGRAM, W R and FISHER, C *Anat Rec*, 1936, 66, 271 *Endocrin*, 1936, 20, 762
- FRAZIER, C H *Arch Neur Psychiat*, 1930, 23, 656
- FRIEDGOOD, H B *Bull Johns Hopkins Hosp*, 1934, 54, 48
- FRY, E G *Endocrin*, 1937, 21, 283
- FULTON, J F, LIDDELL, E G T and RIOCH, D MCK *J Pharm Exp Therap*, 1930, 40, 423
- GEILING, E M K and DELAWDER, A M *Bull Johns Hopkins Hosp*, 1932, 51, 1 *Ibid*, p 335
- GEILING, E M K and OLDHAM, F K *Trans Assoc. Am Phys*, 1937, 52, 132
- GILMAN, A and GOODMAN, L J *Physiol*, 1937, 90, 113
- GRATTAN, J F and JENSEN, H J J *Biol Chem*, 1940, 135, 551
- HAM, A W and HAIST, R E *Am J Path*, 1941, 17, 787
- HANN, F *Von Frankfurt Ztschr f Path*, 1918, 21, 337
- HARRIS, G W *Ciba Found Colloq Endocrinol*, 1952, 4, 106, *Physiol Rev*, 1948, 28, 139
- HARRIS, G W, and assoc *J Physiol*, 1950, III, 347
- HATERIUS, H O and FERGUSON, J K W *Am J Physiol*, 1938, 124, 314
- HELLER, H The excretion of chloride *J Physiol*, 1946, 104, 353 The effect of neurohypophyseal extracts upon the water balance of lower vertebrates, *Biol Rev*, 1945, 20, 147
- HERRING, P T *Quart J Exp Physiol*, 1908, 1, 121
- HERTZ, S and ALBRIGHT, F *J Clin Invest.*, 1934, 13, 698 See also, ALBRIGHT, F and associates *Arch Int Med*, 1934, 54, 315
- HERTZ, S and KRANES, M *Endocrin*, 1934, 18, 350
- HIMSWORTH, H P and SCOTT, D B McV *J Physiol*, 1938, 91, 447
- HOBGEN, L T and WINTON, F R *Proc Roy Soc.*, 1922, B 93, 318
- HOUSSAY, B A, BIASOTTI, A and associates The hypophysis, carbohydrate metabolism and diabetes *Endocrin*, 1931, 15, 511 Phlorrhizin diabetes in fasting or fed hypophysectomized dogs *J Physiol*, 1932, 77, 81 Action de l'extrait antéro-hypophysaire sur le diabètephlorrhizinique *Compt rend Soc biol*, 1933, 112, 497 Action diabétogène de l'extrait antéhypophysaire *Compt rend Soc. biol*, 1932, 111, 479 Hypophyse et métabolisme hydrocarboné *Compt rend Soc biol*, 1922, 86, 115
- HOWELL, W H *J Exp Med*, 1898, 3, 245
- KAMM, O and associates The active principles of the posterior lobe of the pituitary gland I The demonstration of the presence of two active principles II The separation of the two principles and their concentration in the form of potent solid preparations *J Am Chem Soc*, 1928, 50, 573
- KELLER, M D *Proc Soc Exp Med*, 1937, 35, 787
- LI, C H and associates *Science*, 1942, 96, 450, *J Biol Chem* 1943, 149, 413, Harvey Lecture, 1951
- LOEB, L and BASSETT, P B *Proc Soc Exp Biol Med*, 1929, 26, 860
- MAGNUS, R and SCHÄFER, E A *J Physiol*, 1901, 27, ix
- MARIE, P *Acromegaly Brain*, 1889, 12, 59
- MARKS, H P *J Physiol*, 1936, 87, 15 p
- MELVILLE, K I and STEHLE, R L *J Pharm Exp Therap*, 1931, 42, 455
- OLIVER, G and SCHÄFER, E A *J Physiol*, 1895, 18, 277
- PENFIELD, W *Arch Neur Psychiat.*, 1929, 22, 358
- PUTNAM, T J, BENEDICT, E B and TEEL, H M *Am J Physiol*, 1928, 84 157 *Arch Surg*, 1929, 18, 1708
- RAAB, W *Z ges exp Med*, 1926, 49, 179 *Ibid*, 1928, 62, 366
- RASMUSSEN, A The percentage of the different types of cells in the male adult hypophysis *Am J Path*, 1929, 5, 263 The percentage of the different types of cells in the anterior lobe of the adult human female *Ibid*, 1933, 9, 459
- RAWSON, R W *Ann N Y Acad Sc* 1949, 50, 279
- RICHARDSON, K C and YOUNG, F G *J Physiol*, 1937, 91, 352
- RIGHTER, C P *Experimental diabetes insipidus*

- Brain, 1930, 53, 76 The primacy of polyuria in diabetes insipidus A J P, 1935, 112, 481
- RICHTER, C P Experimental diabetes insipidus in relation to the anterior and posterior lobes of the pituitary Am J Physiol, 1934, 110, 124.
- RIDDLE, O Ohio J Science, 1937, 37, 446
- ROWAN, W M Nature, 1925, 116, 494 Ibid, 1927, 119, 351
- ROWLANDS, I W and PARKES, A S J Physiol, 1936, 88, 305
- SAYERS, G and associates J Biol Chem, 1943, 149, 425
- SHIPLEY, R A and LONG, C N H Bioch J, 1938, 32, 2242
- SIMMONDS, M Deut med Woch, 1914, 40, 322
- SMITH, P E Experimental ablation of the hypophysis in the frog embryo Science, 1916, 44, 280 The disabilities caused by hypophysectomy and their repair J Am Med Assoc, 1927, 88, 158 Hypophysectomy and a replacement therapy in the rat Am J Anat, 1930, 45, 205
- SUSMAN, W Brit J Surg, 1935, 22, 539
- UOTILA, U U Endocrin, 1939, 25, 605
- YOUNG, F G Lancet, 1937, 2, 372 Proc Roy Soc. Med, 1938, 31, 1305
- ZONDEK, B J Am Med Assoc, 1935, 104, 637
- Monographs and Reviews*
- ABEL, J J Physiological, chemical and clinical studies on pituitary principles Harvey Lectures, 1923-24, 19, 154
- ABRAMOWITZ, A. A and assoc, Endocrin, 1944, 34, 103
- BURN, J H and DALE, H H Reports on biological standards I Pituitary extracts Med Res Coun, Lond, Spec Rep No 69, 1922
- COLLIP, J B Some recent advances in the physiology of the anterior pituitary J Mount Sinai Hosp, 1934, 1, 28 Antihormones Assoc Res Nerv Ment Dis, 1936, 17, 268 Results of recent studies on anterior pituitary hormones Edin Med J, 1938, 45, 782
- CUSHINO, H. Papers relating to the pituitary body, hypothalamus and parasympathetic nervous system Thomas, Springfield, 1932 "Dyspituitarism" twenty years later, with special consideration of the pituitary adenomas Harvey Lectures, 1932-33, 28, 90 The pituitary body and its disorders Lippincott, Philadelphia, 1912
- VANDYKE, H B Theregulation of water excretion by the neurohypophysis N Y Acad Med, 1953, 29, 24
- EVANS, H M The function of the anterior hypophysis Harvey Lectures, 1923-24, 19, 212 Clinical manifestations of dysfunction of the anterior pituitary J Am Med Assoc., 1935, 104, 464 The growth hormone of the anterior pituitary Ibid, p 1232
- EVANS, H. M., MEYER, K. and SIMPSON, M. E. The growth and gonad stimulating hormones of the anterior hypophysis Memoirs Univ Calif, Vol 11 University of California Press, 1933
- FISHER, C, INGRAM, W R and RANSON, S W Diabetes insipidus Edward Bros, 1938
- GEILING, E M K The pituitary body Physiol Rev, 1926, 6, 62 Some aspects of the comparative anatomy and pharmacology of the pituitary gland Harvey Lectures, 1941-42, 37, 269
- HARRIS, G W Neural control of the pituitary gland, Physiol Rev, 1948, 28, 139
- HOUSSAY, B A Series of lectures on the pituitary New Eng J Med, 1936, 214
- PARKER, G H The reactions of chromatophores Cold Spring Harb Sympos, 1936, 4, 358
- PICKFORD, M Control of the secretion of the antidiuretic hormone of the pituitary Physiol Rev, 1945, 25, 573
- RANSON, S W Some functions of the hypothalamus Harvey Lectures, 1936-37, p 92
- RAWLES, R E Origin of melanophores and their rôle in development of color patterns in vertebrates Physiol Rev, 1948, 28, 383
- REYNOLDS, S R. M Physiology of the uterus Hoeber, N Y, 1939
- RUSSELL, J A The relation of the pituitary to carbohydrate metabolism Physiol Rev, 1938, 18, 1
- SCHARRER, E and SCHARRER, B Neurosecretion Physiol Rev, 1945, 25, 148
- SEVERINGHAUS, A E Cellular changes in the anterior hypophysis with special reference to its secretory activities Physiol Rev, 1937, 17, 556 The pituitary gland Assoc Res Nerv Ment Dis, 1936, 17
- VANDYKE, H B The physiology and pharmacology of the pituitary body Univ Chicago Press, Vol I, 1936, Vol II, 1939
- VERNEY, E B The antidiuretic hormone and the factors which determine its release Proc. Royal Soc Lond Series B, 1947, 135, 25
- WHITE, A Preparation and chemistry of anterior pituitary hormones Physiol Rev, 1946, 26, 574 See also references in Chapter 58

SECTION VII

CHAPTER 58

THE THYROID GLAND

- ABELIN, A Klin Wschr, 1934, 13, 940 Arch Exp Path Pharm, 1934, 173, 146 Ibid, 1934, 175, 151 Ibid, 1936, 181, 250 ABELIN, A and NEFTEL, A Ibid, 1938, 189, 473
- ALBERT, A Endocrin, 1945, 37, 389
- ANDERSON, A B, HARRINGTON, C R and LYON, D M Lancet, 1933, 2, 1081
- ANDRUS, E C Tr Assoc Am Physicians, 1932, 47, 47
- ASTWOOD, E B J Am Med Ass, 1943, 122, 78 J

- Pharm Exp Therap, 1943, 78, 79 Mechanisms of action of antithyroid compounds Ann N Y Acad Sc., 1949, 50, 419 Harvey Lectures, The chemotherapy of hyperthyroidism, 1944-45, 40, 195
- BAUMANN, C Z physiol Chem, 1896, 21, 319 Münch med Woch, 1896, 43, 309 (Quoted by Harrington)
- BOOTHBY, W M and associates The effect of thyroxin on the respiratory and nitrogenous metabolism of normal and myxedematous subjects I A method of studying the reserve or deposit protein with a preliminary report of the results obtained Trans Assoc Am Physicians, 1925, 40, 195
- BRAIN, R. Q J Med, 1938, N S 7, 293 Lancet, 1939, 2, 1217
- CHATIN, M A Series of papers in Compt. rend Acad Sci from 1850, Vol 30, to 1860, Vol 50 See also HERCUS, C E, BENSON, W N and CARTER, C L Endemic goitre in New Zealand and its relation to the soil-iodine J Hygiene, 1925, 24, 321 McCLENDON, J F The distribution of iodine with special reference to goiter Physiol Rev, 1927, 7, 189
- CHESNEY, A M and associated Bull J H. Hosp, 1928, 43, 261
- COINDET, I R Observations on the remarkable effects of iodine in Bronchocele and Scrofula Lond, Longman, Hurst, Reese, Orme and Brown, 1921 (Quoted by Means)
- DEMPSEY, E W Endocrin, 1944, 34, 27 Ann N Y Acad Sc, 1949, 50, 336
- ELMER, A W Endocrin, 1934, 18, 487
- FAGGE, C H Trans Med Chir Soc, London, 1871, 54, 155
- GALLI-MAININI, C Endocrin, 1941, 29, 674
- GOLDSMITH, E D and associates Am J Obst Gynec, 1945, 49, 197
- GUDERNATCH, J F Am J Anat, 1914, 15, 431
- GULL, W W Trans Clin Soc, London, 1874, 7, 180
- HALSTED, W S Johns Hopkins Hosp Rep, 1896, 1, 373
- HARRINGTON, C R Oxidizing agents in the formation of thyroxin in vitro J Chem Soc, 1944, 193 Thyroxine its biosynthesis and its immunochemistry Proc Roy Soc B Lond, 1944, 132, 223
- HARRINGTON, C R and BARGER, G Biochem J, 1927, 21, 169
- HARRINGTON, C R and RANDALL, S S Quart J Pharm, 1929, 2, 501
- HORSLEY, V Proc Roy Soc, 1886, 40, 6
- HUNT, R Am J Physiol., 1923, 63, 257 J Biol Chem, 1905, 1, 33
- HURXTHAL, L M Am Heart J, 1928, 4, 103
- KENDALL, E C Trans Assoc. Am. Physicians, 1915, 30, 420 J Biol Chem, 1919, 39, 125
- KENNEDY, T H and PURVIS, H D Brit J Exp Path, 1941, 22, 241
- KIMBALL, O P Endemic goiter and public health Am J Pub Health, 1928, 18, 587 See also McCURE, R. D Thyroid surgery as affected by the generalized use of iodized salt in an endemic goiter region—preventive surgery Ann Surg, 1934, 100, 924
- LEBLOND, C P Article in Advances in Biol and Med Physics Acad Press Inc, New York, 1948
- LEBLOND, C P and GROSS Endocrin, 1948, 43, 306
- LERMAN, J J Clin Invest, 1940, 19, 555
- LERMAN, J and SALTER, W T Endocrin, 1939, 25, 712
- LUDWIG, W and MUTZENBECHER, P VON Zeit physiol Chem, 1939, 258, 195
- LUNDE, G Chem Rev, 1929, 6, 45
- MACKENZIE, J B and associates Science, 1941, 94, 518 Endocrin, 1943, 32, 185
- MAGNUS-LEVY, A Berl klin Woch, 1895, 32, 650
- MAHER, C C and SITTLER, W W J Am Med Assoc, 1936, 106, 1546
- MARINE, D The rapidity of the involution of active thyroid hyperplasias of brook trout following the use of fresh sea fish as a food J Exp Med, 1914, 19, 376 Remarks on the pathogenesis of Graves' disease Am J Med Sci, 1930, 180, 767 Studies on the etiology of goiter including Graves' disease Ann Int Med, 1930, 4, 423 The thyroid, parathyroids and thymus In Cowdry's Special Cytology, 1932, 2, 797
- MARINE, D and associates The relation of iodine to the structure of the thyroid gland Arch Int Med, 1908, 1, 349 Further observations on the relation of iodine to the structure of the thyroid gland in sheep, dog, hog and ox Ibid, 1909, 3, 66 Relation of iodine to the structure of human thyroids Ibid, 1909, 4, 440 Colloid glands (goitres) their etiology and physiological significance Bull Johns Hopkins Hosp, 1909, 20, 131 The absorption of potassium iodide by perfused thyroid glands and some of the factors modifying it J Pharm Exp Therap, 1915, 7, 557 The absorption of potassium iodide by the thyroid gland *in vivo*, following its intravenous injection in constant amounts Ibid, 1916, 8, 439 The prevention of simple goiter in man J Lab Clin Med, 1917, 3, 40 Arch Int Med, 1920, 25, 661 Influence of glands with internal secretion on the respiratory exchange Am J Physiol, 1921, 57, 135 Further studies in etiology of goiter with particular reference to the action of cyanides Proc Soc Exp Biol Med, 1932, 29, 772 Production of goiter and exophthalmos in rabbits by administration of cyanides Ibid, p 822 976 Further studies in the exophthalmos of rabbits produced by methyl cyanide. Ibid, 1933, 30, 649 Exophthalmos in thyroidectomized guinea pigs by thyrotropic substance of anterior pituitary, and the mechanism involved Ibid, p 910 Exoph-

- thalmos of Graves' disease *Am J Med Sci*, 1934, 188, 565
- MORTON, M E and associates *J Biol Chem*, 1943, 147, 757
- MURRAY, G R *Brit Med J*, 1891, 2, 796
- NAFFZIGER, H C *Arch Ophth*, 1933, 9, I, NAFFZIGER, H C and JONES, O W, *J Am Med Assoc*, 1932, 99, 638
- ORD, W M *Trans Med Chir Soc, London*, 1878, 61, 57
- OSWALD, A Z *physiol Chem*, 1899, 27, 14
- PERKIN, H J, LAHEY, F H and CATTELL, R B *New Eng J Med*, 1936, 214, 45
- PLUMMER, H S The clinical and pathological relationship of simple and exophthalmic goitre *Trans Assoc Am Physicians*, 1913, 28, 587 The function of the thyroid gland *Collected papers, Mayo Clinic*, 1925, 17, 473 The function of the thyroid gland containing adenomatous tissue. *Trans Assoc Am Physicians*, 1928, 43, 159
- POCHIN, E E *Clin Sc*, 1944, 5, 51, 75, *Ciba Found, Colloq Endocrin* 1952, 4, 316
- RAWSON, R W Physiological reactions of thyroid stimulating hormone, its inactivation by exposure to thyroid tissue *in vitro* *Endocrin*, 1942, 30, 240 Physiological reactions of the thyroid hormone *Ann N Y Acad Sc*, 1949, 50, 491 *Recent Prog in Hormone Res*, 1947, 1, 99 *Ciba, Found Colloq Endocrin* 1952, 4, 294
- DE ROBERTIS, E *Anat Rec*, 1941, 80, 219 *Ann N Y Acad Sc*, 1949, 50, 317
- SALTER, The endocrine function of iodine, 1940, Harvard University Press, Cambridge, Mass
- SCHOCKAERT, J A *Proc Socy Exp Biol Med*, 1931, 29, 306
- SEVERINGHAUS, M E Quoted by Lerman, J in Symposium on Glandular Physiology and Therapy *Am Med Assoc Publication*, 1942, p 380
- SMELTZER, G K *Proc Soc Exp Biol Med*, 1936, 35, 128 *Am J Ophth*, 1937, 20, 1189 *Endocrin*, 1938, 21, 1208 *Am J Anat*, 1943, 72, 2 *Anat Rec*, 1943, 85, 3 *Am J Physiol*, 1943, 140, 308 *Ibid*, 1944, 142, 396
- SOFFER, L T and associates *Proc Am Soc Clin Invest*, 1947
- STANLEY, M M and ASTWOOD, E B *Endocrin*, 1947, 41, 66
- SWINGLE, W W *J Exp Zool*, 1919, 27, 397
- WEBSTER, B Studies in the experimental production of simple goiter *Endocrin*, 1932, 16, 617 WEBSTER, B and CHESNEY, A M Endemic goitre in rabbits I Incidence and characteristics *Bull Johns Hopkins Hosp*, 1928, 43, 261 II Heat production in goitrous and non goitrous animals *Ibid*, p 278 III Effects of administration of iodine *Ibid*, p 291
- Monographs and Reviews*
- BRUNTON, C *Exophthalmos Physiol Rev*, 1949, 29, 260
- CARGILL, S L and LESSES, M T Diseases of the thyroid gland *Med Prog, New Eng J Med*, 1946, 235, 717
- DODDS, E C A study of specificity in relation to hormone and other biological relations *Harvey Lectures*, 1934-35, 30, 119 *Glandular Physiology and Therapy*, a Symposium *Am Med Assoc*, 1942
- GOLDSMITH, E D Phylogeny of the thyroid *Ann N Y Acad Sc*, 1949, 50, 283
- HARROW, B and SHERWIN, C P The chemistry of the hormones *The Williams & Williams Co*, Baltimore, 1934
- HOGBen, L T The comparative physiology of internal secretion *Cambridge University Press*, 1927
- KENDALL, E C *Thyroxine Chemical Catalog Co*, New York, 1929
- MANN, I The orbital endocrine problem *Ophthalmic Literature, Lond*, March, 1948
- MARINE, D The present status of the functions of the thyroid gland *Physiol Rev*, 1922, 2, 521 Iodine in the treatment of diseases of the thyroid gland *Medicine*, 1927, 6, 127 The physiology and principal interrelations of the thyroid *J Am Med Assoc*, 1935, 104, 2250 The pathogenesis and prevention of simple or endemic goiter *Ibid*, p 2334
- MCCARRISON, R The etiology of endemic goitre *Bale, London*, 1913
- MCCLENDON, J T Iodine and the incidence of goiter *Univ Minnesota Press*, 1939
- MEANS, J H *The Thyroid Gland*, 2nd Ed *Lippincott Phila*, 1948 *Hyperophthalmopathic Graves disease* *Ann Int Med*, 1945, 23, 779 Thyroid function as disclosed by newer methods of study Series of articles by several authors *Ann N Y Acad Sc*, 1949, 50, 279-508
- MULVANEY, J H The exophthalmos of hyperthyroidism *Am J Ophth*, 1944, 27, 589, 693, 820
- ORR, J B Iodine supply and the incidence of endemic goitre *Med Res Coun, London, Spec Rep No* 154, 1951
- PINCUS, G and THIMANN, K V, Ed *The Hormones Academic Press, N Y*, 1948
- PITT RIVERS, R Mode of Action of Antithyroid Compounds *Physiol Rev*, 1950, 30, 194
- REINEKE, E P The formation of thyroxine in iodinated proteins *Ann N Y Acad Sc*, 1949, 50, 450
- REINHOF, W T, JR A new conception of some morbid changes occurring in diseases of the thyroid gland *Medicine*, 1931, 10, 257
- SALTER, W T The endocrine function of iodine *Harvard Univ Press, Cambridge, Mass* The metabolic circuit of the thyroid hormone *Ann N Y Acad Sc*, 1949, 50, 358

- SCHÄFER, E A SHARPEY The endocrine organs an introduction to the study of internal secretion Longmans, London, 1926
- WOLF, W Endocrinology in modern practice Saunders, Phil, 1937
- WOODS, A C Ocular changes in primary diffuse goiter Medicine, 1946, 25, 113
- ZONDEK, H The diseases of the endocrine glands Arnold, London, 1935

CHAPTER 59

THE ADRENAL GLANDS (SUPRARENAL CAPSULES)

- ADDISON, T London, 1855 Reprinted by the New Sydenham Society, 1868
- ALDRICH, T B Am J Physiol, 1901, 5, 457 J Am Chem Soc, 1905, 27, 1074
- ALLEN, W J and associates J Physiol, 1946, 105, 255
- ALLERS, W D and associates Proc Staff Meet Mayo Clinic, 1935, 10, 406 Ibid, 1936, 11, 283
- BACQ, M M Ann de Physiol Physiochim, 1934, 10, 467
- BAUMANN, E J and KURLAND, S J Biol Chem, 1926, 71, 281
- BEARN, A G and assoc J Physiol, 1951, 115, 430
- BLASCHKO, H J Physiol, 1942, 101, 337
- BOOTHBY, W M and SANDIFORD, I Am J Physiol, 1923, 66, 93
- BRITTON, S W and SILVETTE, H Am J Physiol, 1932, 100, 701, Ibid, 1934, 107, 190
- BROSTER, L R Lancet, 1934, 1, 830 BROSTER, L R and VINES, H W C The adrenal cortex, a surgical and physiological study Lewis, London, 1933
- BROWN-SÉQUARD, C C Compt rend Acad Sci, 1856, 43, 422 (Quoted by Vincent)
- BÜLBING, E J Physiol, 1937, 91, 18P, Brit J Pharmacol, 1949, 4, 234
- BÜLBING, E and BURN, J H J Physiol, 1941, 100, 337, 1949, 109, IIP
- CANNON, W B The emergency function of the adrenal medulla in pain and the major emotions Am J Physiol, 1914, 33, 356 Factors affecting the coagulation time of blood Ibis, 1914, 34, 232 243 251 The isolated heart as an indicator of adrenal secretion induced by pain, asphyxia and excitement Ibid, 1919, 50, 399 Reasons for optimism in the care of the sick New Eng J Med, 1928, 199, 593 Recent studies on chemical mediation of nerve impulses (re sympathin) Endocrin, 1931, 15, 473
- CANNON, W B and associates Emotional stimulation of adrenal secretion Am J Physiol, 1911, 28, 64 Further observations on the denervated heart in relation to adrenal secretion Ibid, 1921, 58, 308 The reflex center for adrenal secretion and its response to excitatory and inhibitory influences Ibid, p 338 Further evidence for reflex and asphyxial secretion of adrenin Ibid, 1922, 61, 215 pseudo-affective medulladrenal secretion Ibid, 1925, 72, 283 The rôle of the adrenal medulla in pseudo-affective hyperglycemia Ibid, p 295 A lasting preparation of the denervated heart for detecting internal secretion, with evidence for accessory accelerator fibers from the thoracic sympathetic chain Ibid, 1926, 77, 326 The influence of motion and emotion on medulladrenal secretion Ibid, 1927, 79, 433 The rôle of adrenal secretion in the chemical control of body temperature Ibid, p 466 The mystery of emotional acceleration of the denervated heart after exclusion of known humoral accelerators Ibid, 1931, 96, 377 A hormone produced by sympathetic action on smooth muscle Ibid, p 392 Sympathin C and sympathin I Ibid, 1933, 104, 557
- CHAIKOFF, I L, quoted by C N H LONG, in discussion, Ciba Found Colloq Endocrinol, 1952, 4, 379
- CHEN, K K J Am Pharm Assoc, 1924, 14, 189, CHEN, K K and SCHMIDT, C F J Am Med Assoc, 1926, 87, 836
- CLEGHORN, R A and associated Am J Physiol, 1939, 126, 466 J Physiol, 1939, 96, 146
- CORI, C F and CORI, G T J Biol Chem, 1928, 79, 309
- DAKIN, H E J Physiol, 1905, 32, xxxiv
- DUNLOP, H A J Physiol, 1929, 67, 349
- ELLIOTT, T R J Physiol, 1912, 44, 374 Ibid, 1914, 49, 38
- VON EULER, N S Acta Med Skand, 1946, 12, 73
- GADDUM, J H and KWIATKOWSKI, H J Physiol, 1938, 94, 87
- GOLDENBERG, M and associates Science, 1949, 109, 534
- GRIFFITH, F R Physiol Rev, 1951, 31, 151
- GROLLMAN, A and FIROR, W M J Biol Chem, 1933, 100, 429
- GUTHRIE, L and EMERY, W D'E Trans Clin Soc, London, 1907, 40, 175
- HARROP, G A The influence of adrenal cortex upon the distribution of body water Bull Johns Hopkins Hosp, 1936, 59, 11 The water and salt hormone of the adrenal cortex Ibid, p 25 HARROP, G A and associates Studies on the suprarenal cortex J Exp Med, 1933, 58, 1 Ibid, 1935, 61, 839
- HARTMAN, F A and KILBORN, L G Am J Physiol, 1918, 45, 111
- HARTMAN, F A and associates A substance which prolongs the life of adrenalectomized cats Proc Soc Exp Biol Med, 1927, 25, 69 The hormone of the adrenal cortex Am J Physiol, 1928, 86, 353 A further study of the hormone of the adrenal cortex Ibid, 1930, 95, 670 A biological method for the assay of cortin Proc Soc Exp Biol Med,

- 1930, 28, 94. Relation of adrenal cortex to development of scurvy. *Proc. Soc. Exp. Biol. Med.*, 1933, 80, 560. A further study of the relation of the adrenal cortex to vitamin C. *Am. J. Physiol.*, 1936, 117, 353.
- HECHTER, O. M., *Trans. 3rd Conf. Adrenal Cortex*, J. Macy Found., 1951, p. 115.
- HENCH, P. S. and associates. *Proc. Staff Meet. Mayo Clin.*, 1949, 24, 181, 277. *Ann. Rheumat. Disease.*, 1949, 8, 90, 97. See also KENDALL, E. C. *Proc. Staff Meet. Mayo Clin.*, 1949, 24, 298.
- HINSWORTH, H. P. and SCOTT, D. B. McN. *J. Physiol.*, 1938, 93, 159.
- HOSKINS, R. G. Relation of the suprarenals to the sympathetic nervous system. *J. Am. Med. Assoc.*, 1927, 88, 2011. The sthenic effect of epinephrin upon intestine. *Am. J. Physiol.*, 1912, 29, 363.
- HOSKINS, R. G. and McCLEURE, C. W. The adrenal glands and blood pressure. *Arch. Int. Med.*, 1912, 10, 343. The relation of the adrenal glands to blood pressure. *Am. J. Physiol.*, 1912, 30, 192.
- KENDALL, E. C. The influence of cortin, insulin and glucose on the metabolism of potassium. *Proc. Staff Meet. Mayo Clinic*, 1938, 13, 519. KENDALL, E. C. and associates. Isolation in crystalline form of the hormone of the adrenal cortex. *Ibid.*, 1932, 9, 245. The chemical nature of the hormone of the adrenal cortex. *Ibid.*, 1936, 11, 351. The structure and physiologic activity of compound B its relation to Reichstein's corticosterone. *Ibid.*, 1937, 12, 136. The chemistry of crystalline substances isolated from the suprarenal gland. *J. Biol. Chem.*, 1936, 114, 613.
- LESCHER, F. G. *Quart. J. Med.*, 1935, N.S. 4, 23.
- LI, C. H. and assoc. *J. Biol. Chem.*, 1943, 149, 413, Harvey Lecture, 1951.
- LOEB, R. F. *Proc. Soc. Exp. Biol. Med.*, 1933, 30, 808. *J. Am. Med. Assoc.*, 1935, 104, 2177. *Bull. N. Y. Acad. Med.*, 1940, 16, 347.
- LOEB, R. F. and associates. Electrolyte balance studies in adrenalectomized dogs, with particular reference to the excretion of sodium. *J. Exp. Med.*, 1933, 57, 775. *Science*, 1939, 90, 496.
- LONG, C. N. H. *Bull. N. Y. Acad. Med.*, 1947, 23, 260. *Proc. Fed. Am. Soc. Exp. Biol.*, 1947, 6, 461.
- LONG, C. N. H. and LUKENS, F. D. W. *J. Exp. Med.*, 1936, 63, 465. LONG, C. N. H., FREY, E. G. and THOMPSON, K. W. *Am. J. Physiol. Proc.*, 1938.
- MENKIN, V. *Am. J. Physiol.*, 1940, 129, 691.
- NICKERSON, M. and GOODMAN, L. S. *Proc. Fed. Am. Soc. Exp. Biol.*, 1948, 7, 397.
- OLIVER, G. and SCHÄFFER, E. A. On the physiological action of extract of the suprarenal capsules. *J. Physiol.*, 1894, 16, 1. The physiological effects of extracts of the suprarenal capsules. *Ibid.*, 1895, 18, 230.
- OPSAHL, J. C., *Yale J. Biol. Med.*, 1949, 21, 255, 433, 487, *Trans., 2nd Confer. Adrenal Cortex*, 1950.
- REDFIELD, A. C. *J. Exp. Zool.*, 1918, 26, 275.
- REICHSTEIN and associates. *Helv. chim. Acta.*, 1934, 19, 1107. *Nature*, 1937, 139, 26, 925.
- ROGOFF, J. M. Clinical and experimental studies on adrenal insufficiency and Addison's disease and the treatment of such conditions by interrenalin. *Proc. Calif. Acad. Med.*, 1930, 1, 7. Diagnosis and treatment of Addison's disease. *Can. Med. Assoc. J.*, 1931, 24, 43.
- ROWNTREE, L. G. and BALL, R. G. *Endocrin.*, 1933, 17, 263.
- SAYERS, G., See *Monographs and Reviews*.
- SMITHWICK, R. H., FREEMAN, N. E. and WHITE, J. C. *Arch. Surg.*, 1934, 29, 759.
- SOFFER, L. J. *Bull. N. Y. Acad. Med.*, 1947, 23, 479.
- SOSKIN, A. *Am. J. Physiol.*, 1927, 83, 162.
- STEWART, G. N. and ROGOFF, J. M. The influence of adrenal extracts on the survival period of adrenalectomized dogs. *Science*, 1927, 66, 327. *Am. J. Physiol.*, 1928, 84, 660. *Ibid.*, 1929, 91, 254. The influence of "heat" on the survival period of dogs after adrenalectomy. *Ibid.*, 1928, 86, 20. The survival period of untreated adrenalectomized cats. *Ibid.*, 1929, 88, 162. Suprarenal cortical extracts in suprarenal insufficiency (Addison's disease). *J. Am. Med. Assoc.*, 1929, 92, 1569.
- STOLTZ, F. *Ber. Deut. Chem. Ges.*, 1904, 37, 4149.
- SWINGLE, W. W. and associates. An aqueous extract of the suprarenal cortex which maintains the life of bilaterally adrenalectomized cats. *Science*, 1930, 71, 321. 489. *Am. J. Physiol.*, 1931, 96, 164. The effect of a lipid fraction upon the life-span of adrenalectomized cats. *Ibid.*, p. 153. Revival of comatose adrenalectomized cats with an extract of the suprarenal cortex. *Science*, 1930, 72, 75. 483. Further purification of the adrenal cortical hormone. *Proc. Soc. Exp. Biol. Med.*, 1932, 29, 1267. Metabolic studies in Addison's disease. *Am. J. Med. Sci.*, 1932, 183, 1. Studies in the suprarenal cortex. *J. Exp. Med.*, 1933, 58, 17. The function of the adrenal cortical hormone and the cause of death from adrenal insufficiency. *Science*, 1933, 77, 58. The effect of hemorrhages on the normal and adrenalectomized dog. *Am. J. Physiol.*, 1934, 107, 259. The effect of sodium chloride administration upon adrenalectomized dogs not receiving extract. *Ibid.*, 1934, 108, 159.
- SCHWARTZMAN, G., *Proc. Soc. Exper. Biol. Med.*, 1952, 79, 573.
- SZENT-GYÖRGYI, A. *Am. J. Physiol.*, 1929, 90, 536. *Science*, 1930, 72, 125.
- TAKAMINE, J. *Am. J. Pharm.*, 1901, 73, 523 (Quoted by Vincent.)
- THORN, G. W. and associates. *Proc. Soc. Exp. Biol. Med.*, 1936, 35, 247.
- VULPIAN, E. F. A. *Compt. rend. Acad. Sci.*, 1856, 43, 663.

- WEAVER, H M and NELSON, W O *Anat Rec*, 1943, 85, 51
- WELCH, A D *Am J Physiol*, 1934, 108, 691
- WILDER, R M and associates *J Am Med Assoc*, 1938, 111, 117
- Monographs and Reviews*
- BEYER, K H *Sympathomimetic amines, the relation of structure to their action and inactivation* *Physiol Rev*, 1946, 26, 169
- BRITTON, S W *Adrenal insufficiency and related considerations* *Physiol Rev*, 1930, 10, 617
- BROSTER, L R et al *The adrenal cortex and intersexuality* Chapman, Lond, 1938
- BURN, J H *Adrenaline and acetylcholine in the nervous system* *Physiol Rev*, 1945, 25, 377
- CANNON, W B *Bodily changes in pain, hunger, fear and rage* Appleton, New York, 1915
- CANNON, W B and ROSENBLUETH, A *Autonomic neuro-effector systems* Macmillan, New York, 1937
- GROLLMAN, A *The adrenals* The Williams & Wilkins Co, Baltimore, 1936
- HOSKINS, R G *The relation of the adrenals to the circulation* *Physiol Rev*, 1922, 2, 343
- LERNER, A B and FITZPATRICK, T B *Biochemistry of Melanin Formation* *Physiol Rev*, 1950, 30, 91
- LOEB, R F *The adrenal cortex and electrolyte behavior* Harvey Lectures, 1941-42, 37, 100
- LONG, C N H *The influence of the pituitary and adrenals upon carbohydrate metabolism* *Medicine*, 1937, 16, 215
- NICKERSON, M *Pharmacology of adrenergic blockade* *Pharmacol Rev*, 1949, 1, 27
- PARKES, A G *Physiol Rev*, 1951, 31, 18P
- SAYERS, G *The adrenal cortex and homeostasis* *Physiol Rev*, 1950, 30, 241, *The pituitary-adrenal system* *Rec Prog Hormone Research*, 1948, 2, 81
- STEWART, G N *Adrenalectomy and the relation of the adrenal bodies to metabolism* *Physiol Rev*, 1924, 4, 163 *The adrenal glands* *Arch Int Med*, 1929, 43, 753
- SWINGLE, W W and PFIFFNER, J J *The adrenal cortical hormone* *Medicine*, 1932, 11, 371
- See also references in Chapter 58
- BINGER, C *Toxicity of phosphates in relation to blood calcium and tetany* *J Pharm Exp Therap*, 1917, 10, 105
- BODANSKY, A and JAFFE, H L *Parathormone dosage and serum calcium and phosphorus in experimental chronic hyperparathyroidism leading to ostitis fibrosa* *J Exp Med*, 1931, 53, 591 JAFFE, H L and BODANSKY, A *Experimental fibrous osteodystrophy (ostitis fibrosa) in hyperparathyroid dogs* *Ibid*, 1930, 52, 669 BODANSKY, A, BLAIR, J E and JAFFE, H L *Experimental hyperparathyroidism in guinea pigs leading to osteitis fibrosa* *J Biol Chem*, 1930, 88, 629 JAFFE, H L, BODANSKY, M and BLAIR, J E *The effects of parathormone and ammonium chloride on the bones of rabbits* *J Exp Med*, 1932, 55, 695
- BREITER, H and associates *J Nutrit*, 1941, 21, 351
- CAMERON, A T and MOORHOUSE, V H K *J Biol Chem*, 1925, 63, 687 *Trans Roy Soc Can*, 1925, 19, Sec V, 39
- COLLIP, J B *The extraction of a parathyroid hormone which will prevent or control parathyroid tetany and which regulates the level of blood calcium* *J Biol Chem*, 1925, 63, 395 COLLIP, J B and CLARK, E P *Further studies on the physiological action of a parathyroid hormone* *Ibid*, 1925, 64, 485 COLLIP, J B, CLARK, E P and SCOTT, J W *The effect of a parathyroid hormone on normal animals* *Ibid*, 1925, 63, 439
- ELLSWORTH, R J *Clin Invest*, 1932, 11, 1011
- GLEY, E *Arch de Physiol*, 1893, 5 Série, 5, 766
- GREENBERG, D M and associates *Am J Physiol*, 1942, 137, 459
- GREENWALD, I J *Biol Chem*, 1924, 61, 33
- HASTINGS, A B, ROSEBERRY, H H and MORSE, J K *J Biol Chem*, 1931, 90, 395
- HIGGINS, G M and SHEARD, C *Am J Physiol*, 1928, 85, 299
- HOLTZ, F *Klin Wchnschr*, 1934, 13, 104
- HUGGINS, C B *Arch Surg*, 1931, 22, 377
- HUNTER, D *Brit J Surg*, 1931, 19, 203
- IMRIE, C G and JENKINSON, C N *J Physiol*, 1933, 79, 218
- KAY, H D *J Biol Chem*, 1930, 89, 249
- KAY, H D and GRAHAM, W R *J Dairy Res*, 1933, 5, 54
- LOEB, J *On the different effect of ions upon myogenic and neurogenic rhythmical contractions and upon embryonic and muscular tissue* *Am J Physiol*, 1900, 3, 383 *On an apparently new form of muscular irritability (contact irritability?) produced by solutions of salts (preferably sodium salts) whose anions are liable to form insoluble calcium compounds* *Ibid*, 1901, 5, 362 *Calcium in permeability and irritability* *J Biol Chem*, 1915, 23, 423

CHAPTER 60

THE PARATHYROID GLANDS AND CALCIUM METABOLISM

- ALBRIGHT, F, DRAKE, T G and SULKOWITCH, H W *Bull J Hopk Hosp*, 1937, 60, 377
- ANDERSON, I A and LYALL, A *Quart J Med*, 1939, 8, 209
- ANDERSON, W A D *Endoc*, 1939, 24, 372
- BAUER, W, AUB, J C and ALBRIGHT, F *J Exp Med*, 1929, 49, 145

- MACCALLUM, W G and VOEGTLIN, C J *Exp Med*, 1909, 11, 118
- MCLEAN, F C and HASTINGS, A B J *Biol Chem*, 1934, 107, 337 *Ibid*, 1935, 108, 285
- MACLEOD, J J R and TAYLOR, N B *Trans Roy Soc Can*, 1925, 19, Sec V, 27
- MANDL, F *Zentralblatt für Chirurgie* 1926, 53, 260 (Quoted by Hunter) See also, SNAPPER, I Parathyroid tumor and changes of the bones *Arch. Int Med*, 1930, 46, 506 TAYLOR, H *Osteitis fibrosa*, an experimental study *Brit. J Surg*, 1934, 22, 561
- PAPPENHEIMER, A M J *Exp Med*, 1936, 64, 965
- PATON, D N and FINDLAY, J *Quart J Exp Physiol*, 1916, 10, 203 233 315 377
- PATT, H. M and LUCKHARDT, A. B *Endocrin*, 1942, 31, 384
- ROBISON, R *Biochem J*, 1923, 17, 286
- ROBISON, R and SOAMES, K M Part II *Biochem J*, 1924, 18, 740 *Ibid*, 1925, 19, 153
- SCHOUR, I and HAM, A *Arch Path*, 1934, 17, 22
- SELYE, H On the stimulation of new bone formation with parathyroid extract and irradiated ergosterol *Endocrin*, 1932, 16, 547 Action of parathyroid hormone on the epiphyseal junction of the young rat. *Arch Path*, 1932, 14, 60 Condition simulating human scleroderma in rats injected with parathyroid hormone *J Am Med Assoc*, 1932, 99, 108 PUGSLEY, L I and SELYE, H. The histological changes in the bone responsible for the action of parathyroid hormone on the calcium metabolism of the rat *J Physiol*, 1933, 79, 113
- SHERMAN, H. C and HAWLEY, E J *Biol Chem*, 1922, 53, 375
- SHIPLEY, P G, KRAMER, B and HOWLAND, J *Biochem J*, 1926, 20, 379 SHIPLEY, P G and HOLT, L E *Bull Johns Hopkins Hosp*, 1927, 40, 1 *Ibid*, 41, 437
- TAYLOR, N W and SHEARD, C J *Biol Chem*, 1929, 81, 479
- TAYLOR, N D, WELD, C B, BRANION, H D and KAY, H. D *Can Med Assoc. J*, 1931, 24, 763 *Ibid*, 1931, 25, 20
- THOMSON, D L and PUGSLEY, L I *Am J Physiol*, 1932, 102, 350
- WATT, J C *Arch Surg*, 1925, 10, 983
- WEST, R *Brain*, 1935, 58, 1
- BARR, D P Pathological calcification *Physiol Rev*, 1932, 12, 593
- COLLIP, J B The parathyroid glands *Medicine*, 1926, 5, 1
- DRAGSTEDT, L R. The physiology of the parathyroid glands *Physiol Rev*, 1927, 7, 499
- HAM, A Cartilage and bone. In Cowdry's *Special Cytology*, 1932, 2, 981
- HAM, A The last hundred years in the study of bone. *J Am Dent Assoc.*, 1934, 21, 3
- HARRISON, H E and HARRISON, H C *J Clin Invest*, 1941, 20, 47
- HESS, A F Rickets, including osteomalacia and tetany Lea, Philadelphia, 1929
- HUGGINS, C Composition of bone and the function of the bone cell *Phys Rev*, 1937, 17, 119
- HUNTER, D The significance to clinical medicine of studies in calcium and phosphorus metabolism *Lancet*, 1930, 1, 897 947 999
- HUNTER, D and AUB, J C Lead studies XV The effect of the parathyroid hormone on the excretion of lead and of calcium in patients suffering from lead poisoning *Quart J Med*, 1926, 20, 123
- KAY, H D Phosphatase in growth and disease of bone *Physiol Rev*, 1932, 12, 384
- LOGAN, M A Recent advances in the chemistry of calcification *Physiol Rev*, 1940, 20, 522
- MACCALLUM, W G On the pathogenesis of tetany *Medicine*, 1924, 3, 137
- MACLEAN, F C Physiology of bone. *Ann Rev Physiol*, 1943, 5, 79
- MACLEAN, F C and BLOOM, W *Science*, 1937, 85, 24
- ROBISON, R The significance of phosphoric esters in metabolism New York University Press, 1932
- SCHMIDT, C L A. and GREENBERG, D M Occurrence, transport and regulation of calcium, magnesium and phosphorus in the animal organism *Physiol Rev*, 1935, 15, 297
- THOMPSON, D L and COLLIP, J B The parathyroid glands *Physiol Rev*, 1932, 12, 309
- See also references in Chapter LVIII

CHAPTER 61

THE ENDOCRINE ORGANS OF SEX. THE THYMUS AND PINEAL GLANDS

Monographs and Reviews

- ALBRIGHT, F and REIFENSTEIN, E C The parathyroid glands and metabolic bone disease The Wilhams and Wilkins Co., Baltimore, 1948
- AUB, J C Calcium and phosphorus metabolism Harvey Lectures, 1928-29, 24, 151
- AUB, J C and associates Lead poisoning *Medicine*, 1925, 4, 1
- BARNICOR, N A *Nature*, 1948, 162, 848

- ANDERSEN, D H, J *Physiol*, 1935, 85, 162
- ALLAN, H and WILES, P J *Physiol*, 1932, 76, 23
- ALLEN, E and DOISY, E A An ovarian hormone, preliminary report on its localization, extraction and partial purification and action in test animals *J Am Med Assoc*, 1923, 81, 819 The induction of a sexually mature condition in immature females by injection of the ovarian follicular hormone *Am J Physiol*, 1924, 69, 577 ALLEN, E, PRATT, J P and DOISY, E A The ovarian follicular hormone, its distribution in the human genital tissues *J Am Med Assoc*, 1925, 85, 399